PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/54, C12Q 1/68, A61K 48/00, C12N 15/86, 15/67

A2

(11) International Publication Number: **WO 97/04104**

(43) International Publication Date: 6 February 1997 (06.02.97)

(21) International Application Number: PCT/US96/11673

(22) International Filing Date: 12 July 1996 (12.07.96)

(30) Priority Data:
60/001,199 14 July 1995 (14.07.95) US

(60) Parent Application or Grant
(63) Related by Continuation
US 60/001,199 (CIP)
Filed on 14 July 1995 (14.07.95)

(71) Applicant *(for all designated States except US)*: THE JOHNS HOPKINS UNIVERSITY [US/US]; School of Medicine, Suite 2-100, 2024 East Monument Street, Baltimore, MD 21205 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: PEDERSEN, Peter, Lynn [US/US]; 9414 Diamond Back Drive, Columbia, MD 21045 (US). MATHUPALA, Saroj, P. [LK/US]; Apartment B, 6510 Falkirk Road, Baltimore, MD 21239 (US). REMPEL, Annette [US/US]; Apartment 1229, 116 West University Parkway, Baltimore, MD 21210 (US).

(74) Agents: KAGAN, Sarah, A. et al.; Banner & Allegretti, Ltd., Suite 1100, 1001 G Street, N.W., Washington, DC 20001 (US).

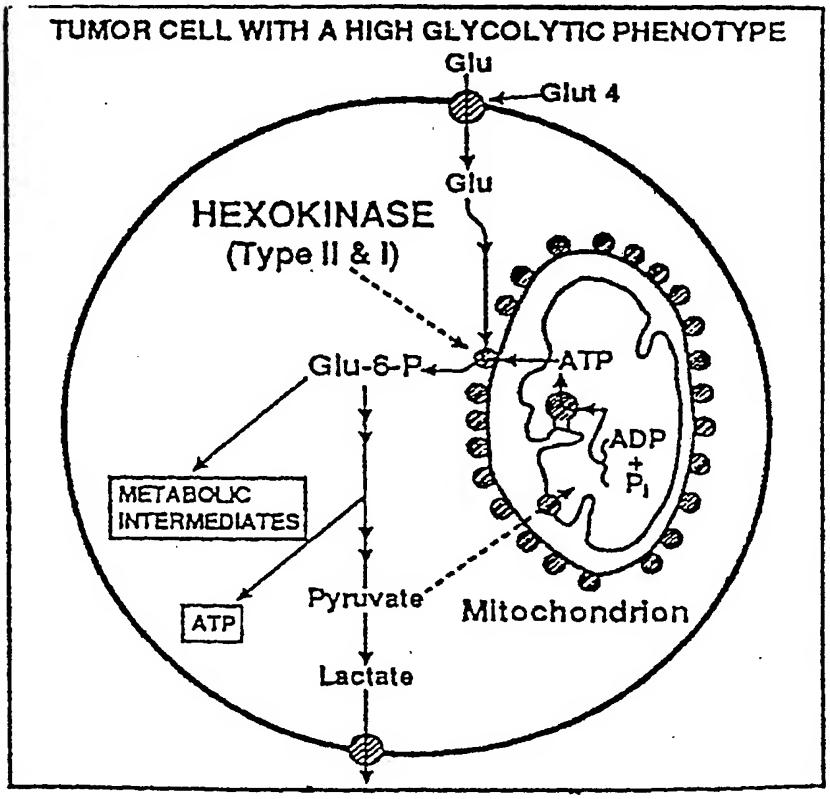
(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: TUMOR TYPE II HEXOKINASE TRANSCRIPTION REGULATORY REGIONS

(57) Abstract

The present invention relates to a tumor promoter involved in the regulation of glucose catabolism in neoplastic tissues. In particular, this promoter region contains numerous response elements that are involved in regulation of transcription of the Type II hexokinase gene in tumor cells. Such elements are of value for diagnostic and therapeutic applications, such as in controlling tumor growth. In addition, the entire promoter region (about 4.3 kbp) or regulatory segments (response elements) contained therein may be used for expression of naturally-occurring or foreign proteins. Such proteins may be derived from mammalian cells and expressed under the control of transcription factors that bind to specific response elements within the tumor Type II hexokinase promoter. Finally, the promoter is useful in gene therapeutic approaches to diseases including diabetes and cancer.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

TUMOR TYPE II HEXOKINASE TRANSCRIPTION REGULATORY REGIONS

The invention was made with government support under NIH Grant CA 32742. The government may have certain rights in this invention.

TECHNICAL FIELD OF THE INVENTION

The present invention relates to a tumor promoter and transcription regulatory region involved in the regulation of glucose catabolism in neoplastic tissues. In particular, this transcription regulatory region contains numerous response elements that are involved in regulation of transcription of the Type II hexokinase gene in tumor cells.

BACKGROUND OF THE INVENTION

It has been known for over six decades that one of the most consistent and profound biochemical phenotypes of cancer cells is their increased rate of glucose utilization. In recent years, hexokinase, which catalyzes the first step of the glycolytic pathway and which is highly overexpressed in tumor cells, has been shown to be a major player in this process of glucose catabolism in cancer cells. In comparison to normal cells, rapidly-growing tumors have elevated hexokinase activity levels (Parry, D. and Pedersen, P., (1983), *J. Biol. Chem.*, 258:10904-10912). Glycolysis is defined as the metabolism of glucose to yield either lactic acid under anaerobic conditions or pyruvate under aerobic conditions, the latter being further metabolized to carbon dioxide and water.

Rapidly-growing cancer cells have the ability to maintain an increased rate of glucose utilization and the capacity to sustain high rates of glycolysis under aerobic conditions (Warburg, O., (1930), *The Metabolism of Tumors*, Arnold Constable, London); Weinhouse, S., (1966), *Gann Monogr.*, 1:99-115; Bustamante, E. and Pedersen, P., (1977), *Proc. Natl. Acad. Sci USA*, 74:3735-3739; Aisenberg, A., (1961), In: *The Glycolysis and Respiration of Tumors*, Academic Press, London; Pedersen, P., (1978), *Prog. Exp. Tumor Res.*, 22:190-274). This elevated rate of glucose catabolism is important for highly malignant tumors, such as tumors derived from liver, kidney, and brain, which may obtain over 50% of their energy, and the

anabolic precursors for biosynthetic pathways, via glycolysis (Nakashima, et al., (1984), *Cancer Research*, 44:5702-5706; Bustamante, et al., (1981), *J. Biol. Chem.*, 256:8699-8704).

There are four hexokinase isozymes in mammals designated as Type I-IV. Type IV is also called glucokinase. In general, hexokinases, known as D-hexose 6-phosphotransferases, catalyze the following reaction:

$$\text{Glucose} + \text{MgATP} \underset{<-----}{\overset{\text{----------}>}{}} \text{Glucose-6-Phosphate} + \text{MgADP}.$$

Isozymes are multiple forms of a given enzyme that may occur within a single species of organism or within a single cell. Such multiple forms can be detected and separated by gel electrophoresis of cell extracts; as they are coded by different genes, they differ in amino acid composition and thus, in their isoelectric point values (Lehninger, (1975), In: Biochemistry, Worth Publishers, Inc., NY, pg. 244). (Wilson, J.E., (1985), In: Regulation of Carbohydrate Metabolism (Breitner, R., ed.) CRC Press I, pgs. 45-86; Rijksen, et al., (1985), In: Regulation of Carbohydrate Metabolism (Breitner, R., ed.) CRC Press I, pgs. 89-99; Pilkis, et al., (1994), *J. Biol. Chem.*, 269:21925-21928). Types I-III exhibit very low Kms (0.02-0.13 mM) for glucose (high affinities), are product inhibited by glucose-6-phosphate, (Glu-6-P), and have a molecular mass near 100 kDa. The Type IV isozyme, in contrast to Types I-III, has a high Km (5-8 mM) for glucose, is insensitive to Glu-6-P inhibition, and has a mass near 50kDa.

The distribution of the four isozymes is tissue specific. Type I is found normally in brain and kidneys, Type II is found in skeletal muscle and adipose tissue, Type III is found in low amounts in several tissues, and Type IV is found predominantly in the liver and pancreas. Within normal liver cells, Type IV hexokinase is the predominantly expressed isoform (Printz, et al., (1993), Annu. Rev. Nutr., 13:463-496), and transcription of this enzyme is enhanced by both glucose and insulin (fed state), and inhibited by glucagon (fasted state) (Granner and Pilkis, (1990),

J. Biol. Chem., 265:10173-10176). Significantly, the Type II hexokinase gene, which is markedly overexpressed in hepatoma cells, is essentially silent in liver.

In comparison to normal cells, the activity of Type II hexokinase is markedly elevated in highly glycolytic, rapidly growing tumors (Pedersen, P.L., (1978), Prog. Exp. Tumor Res., 22:190-274; Bustamante, supra; Arora, et al., (1988), *J. Biol. Chem.*, 263:14422-14428; Parry, supra). Two factors are known to be largely responsible for this enhanced activity, one of which involves a propensity for the tumor enzyme to bind to the outer mitochondrial membrane, and the other which involves the enzyme's overproduction. Mitochondrial binding of tumor hexokinase to the outer membrane has been intensely studied (Bustamante, supra; Rose, et al., (1967), *J. Biol. Chem.*, 242:1635-1645; Parry and Pedersen, (1983), J. Biol. Chem., 258:10904-10912). Mitochondrial binding provides the enzyme with preferential access to mitochondrially-generated ATP and increases the activity and stability of the enzyme. (Arora, *supra*). Mitochondrial membrane binding also reduces the sensitivity to product inhibition by G-6-P, which is an important regulator of hexokinase in normal cells (Bustamante, supra; Rose, supra; Gumaa, et al., *Biochem. Biophys. Res. Comm.*, 36:771-779; Kurokawa, et al., (1981), *Biochem. Int.*, 2:645-650; Inui, et al., (1979), *J. Biochem.*, 85:1151-1156). The end product of the hexokinase reaction, glucose-6-phosphate (G-6-P), serves not only as a source of ATP via glycolysis but is also a key intermediate in metabolic processes essential for cell growth and proliferation, as shown in Figure 1.

The cDNAs representing each of the four isozymes have been cloned and sequenced (Nishi, et al., (1988), *Biochem. Biophys. Res. Comm.*, 157:937-943; Schwab, et al., (1989), *Proc. Natl. Acad. Sci. USA*, 86:2563-2567; Thelen, et al., (1991), *Arch. Biochem. Biophys.*, 286:645-651; Schwab, et al., (1991), *Arch. Biochem. Biophys.*, 285:354-370; Andreone, et al., (1989), *J. Biol. Chem.*, 264:363-369). In addition, hexokinase cDNAs have been cloned and isolated from different tumors (Thelen, supra; Arora, et al., (1990), *J. Biol. Chem.*, 265:6481-6488).

Using a Type I hexokinase cDNA probe (from brain), the hexokinase isozyme expressed in a mouse hepatoma cell line (c37) has been cloned and characterized

(Arora, *supra*) and shown to be approximately 92% identical to the hexokinase I sequence derived from rat brain (Schwab, et al., (1989), Proc. Natl. Acad. Sci. USA, 86:2563-2567) and human kidney (Nishi, et al., (1988), Biochem. Biophys. Res. Comm., 157:937-943). Now recognized as Type I, it is overexpressed in the AS-30D hepatoma cells (Figure 1), but to a much lesser extent than Type II hexokinase. Thus, regulatory regions of the Type I and Type II hexokinase forms may share some common activating elements. Nevertheless, in those rapidly growing, highly glycolytic cancers studied to date (Shinohara, supra; Nakashima, supra; Thelen, supra; Sato, et al., (1972), Gann Monograph. on Cancer Res., 13:279-288; Kikuchi, et al., (1972), Cancer, 30:444-447; Hammond and Balinsky, (1978), Cancer Res., 38:1323-1328; Singh, et al., (1978), J. Cell Physiol., 97:285-292; Rose and Warms, (1982), Arch. Biochem. Biophys., 213:625-634; Rempel, et al., (1994), Biochem. Biophys. Acta., 219:660-668), it is the Type II hexokinase isozyme that either dominates or increases upon transformation. Thus, Type I, although elevated somewhat in highly glycolytic tumors, is elevated to a much lesser extent than Type II (Nakashima, et al., (1988), *Cancer Res.*, 48:913-919; Rempel, et al., (1994) *Biochem. Biophys. Acta.*, 1219:660-668; Parry, et al., (1983), *J. Biol. Chem.*, 258:10904-10912; Kurokawa, et al., (1982), *Mol. Cell. Biochem.*, 45:151-157).

Both Type I and Type II isoforms bind to the outer mitochondrial membrane pore protein, which is known as VDAC (Nakashima, et al., (1986), *Biochem.*, 25:1015-1021). This binding markedly reduces the enzymes' sensitivity to product inhibition by Glu-6-P (Bustamante, et al., (1977), *Proc. Natl. Acad. Sci. USA*, 74:3735-3739), provides preferred access to mitochondrially-generated ATP (Arora, et al., (1988), *J. Biol. Chem.*, 263:14422-14428), and also provides protection against proteolytic degradation (Rose, et al., (1982), *Arch. Biochem. Biophys.*, 213:625-634). In addition to these properties and the high content of the tumor enzyme (100-fold elevation), Glu-6-P is rapidly produced. Glu-6-P is a key metabolic intermediate serving as a major carbon source for most biosynthetic products essential for cell growth and division, but also results in ATP synthesis during its catabolism to lactic acid (Figure 1). It is estimated that under aerobic conditions more than half the ATP

produced in some tumor cells is derived from glycolytic reactions (Aisenberg, A.C., (1961), In: *The Glycolysis and Respiration of Tumors*, Academic Press, London, pgs. 8-11; Nakashima, et al., (1984), *Cancer Res.*, 44:5702-5706), in contrast to normal cells where this value is usually less than 10 percent. Under hypoxic (low oxygen) or anaerobic conditions, the already high glycolytic rate may double (Weinhouse, S., (1972), *Cancer Res.*, 32:2007-2016), allowing tumor cells to thrive while neighboring normal cells become growth deficient.

The marked overexpression of Type II hexokinase in tumor cells characterized phenotypically by a high glucose catabolic rate is consistent with an enhanced rate of transcription. Elevated mRNA levels for hexokinase have been found in all highly glycolytic tumor lines examined to date (Shinohara, FEBS Lett, supra; Rempel, supra; Mathupala, et al, (1995), *J. Biol. Chem.*, 16918-16925).

Mitochondrial-bound hexokinase is also found in some normal tissues, including brain and skeletal muscle (Wilson., J., (1985), In: *Regulation of Carbohydrate Metabolism*, Beitner, R, ed., 1:45-85, CRC Press, Inc., Boca Raton, Florida), but in significantly lower amounts than in rapidly-growing cancer cells. In addition, mRNA is markedly overexpressed in the AS-30D hepatoma cell line relative to normal liver and skeletal muscle. The deduced amino acid sequence for the hepatoma Type II enzyme differs in only 4 amino acids from the skeletal muscle Type II enzyme (Thelen, supra). Moreover, regions of the enzyme predicted to be involved in catalysis and Glu-6-P inhibition are nearly identical to those found in the tumor Type I isozyme (Mathupala, Rempel, and Pedersen, unpublished results). As the Type II hexokinase isozyme of normal and tumor tissues are essentially identical, the contribution of the Type II enzyme to the high glycolytic phenotype of many tumor cells is mainly due to elevated transcription and not to a property of the enzyme itself. Thus, the role of hexokinase, (ATP: D-hexose 6-phosphotransferase; E.C. 2.7.1.1), which commits glucose to catabolism in the

first step of the glycolytic pathway, in particular the Type II isoform has come under increased scrutiny in efforts to understand the molecular basis for the aberrant glycolytic phenotype (Bustamante, supra; Nakashima, et al., (1986), *Biochemistry*, 25:1015-1021; Arora, K. and Pedersen, P., (1988), *J. Biol. Chem.*, 263:17422-17428). Significantly, hexokinase is a potential target for arresting tumor cell growth (Floridi, et al., (1981), *J. Natl. Cancer Institute*, 66:497-499; Floridi, et al., (1981), *Cancer Res.*, 41:4661-4666). The role of hexokinases, particularly Type II and Type IV, in diabetes is now attracting considerable attention as well (Pilkis, supra).

There remains a need in the art for methods and drugs to effectively treat both neoplasia and diabetes.

SUMMARY OF THE INVENTION

It is an object of the invention to provide a DNA fragment useful in the regulation of transcription of a downstream open reading frame.

It is another object of the invention to provide a method of screening for potential drugs useful in the treatment of cancers and diabetes.

It is yet another object of the invention to provide a method of treating cells which are neoplastic.

It is still another object of the invention to provide nucleic acid probes useful in the diagnosis of cancers.

It is a further object of the invention to provide methods for diagnosing tumors.

It is yet another object of the invention to provide a vector for regulated expression of desired proteins in a mammalian cell.

It is still another object of the invention to provide a method for increasing glycolysis in cells.

These and other objects of the invention are provided by one or more of the embodiments shown below. In one embodiment of the invention an isolated hexokinase II DNA fragment capable of regulating transcription of a downstream open reading frame is provided. The fragment comprises at least one of the response elements identified in Figure 11.

According to another embodiment of the invention a method of screening for potential drugs which affect with regulated transcription of tumor hexokinase II is provided. The method comprises the steps of:

contacting a test substance with a reporter gene fusion comprising an isolated hexokinase II DNA fragment capable of regulating transcription of a downstream open reading frame, wherein the fragment comprises at least one of the response elements identified in Figure 11; and

measuring transcription of the reporter gene in the presence of the test substance;

identifying a test substance as a potential drug which increases or decreases the transcription of the reporter gene.

In yet another embodiment of the invention a method of treating cells which overexpress hexokinase II is provided. The method comprises the step of:

administering to cells which overexpress hexokinase II a gene fusion comprising a toxic gene and a an isolated hexokinase II DNA fragment capable of regulating transcription of a downstream open reading frame, wherein the fragment comprises at least one of the response elements identified in Figure 11, whereby the toxic gene is expressed in the cells.

According to still another embodiment of the invention a n isolated nucleic acid probe is provided. The probe comprises at least 15 contiguous nucleotides selected from the sequence of SEQ ID NO: 1.

In another aspect of the invention a method for diagnosing tumors which overexpress hexokinase is provided. The method comprises the steps of:

determining copy number of a hexokinase II gene in a tissue sample suspected of being neoplastic;

wherein a determined copy number of greater than two indicates neoplasia.

According to another aspect of the invention a method for diagnosing neoplastic tissues is provided. The method comprises the step of:

determining whether cells in a tissue sample suspected of being neoplastic contain a hexokinase II gene which is unmethylated, an unmethylated hexokinase II gene indicating neoplasia.

In another aspect of the invention a vector for expression of a desired protein in a mammalian cell is provided. The vector comprises: an isolated hexokinase II DNA fragment capable of regulating transcription of a downstream open reading frame, wherein the fragment comprises at least one of the response elements identified in Figure 11.

In another embodiment of the invention a method is provided for increasing glycoslysis in cells. The method comprises the step of:

introducing into cells an unmethylated DNA molecule comprising:

a hexokinase II DNA fragment capable of regulating transcription of a downstream open reading frame, wherein the fragment comprises at least one of the response elements identified in Figure 11; and

a nucleic acid encoding a hexokinase II, wherein the hexokinase II DNA fragment is covalently and operatively linked to the nucleic acid encoding a hexokinase II.

The present invention thus provides the art with methods and reagents for treating diseases such as cancer and diabetes, as well as methods for diagnosing neoplasia. In addition, the method provides methods for screening for new therapeutic agents for such diseases.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a schematic drawing of the basic metabolic pathway for glucose in tumor cells with a high glycolytic phenotype.

Figure 2 shows a Northern blot analysis carried out on total RNA to determine the expression levels of Type I and Type II hexokinase in AS-30D hepatoma cells relative to their expression in normal rat liver. Specifically, for Type I hexokinase, the blot was hybridized with a 0.7 kbp fragment of the c37 tumor Type I hexokinase cDNA (Panel 2A) and for Type II hexokinase, a full length cDNA of rat skeletal muscle Type II hexokinase was used (Panel 2B). Single mRNA species in AS-30D cells (lane 4) were detected similar in size to the hybridization bands obtained with rat brain (lane 1) and skeletal muscle RNA (lane 2), which were used as controls for the Type I and Type II hexokinase messages, respectively. The Type I and Type II hexokinase messages were below the detection level in normal rat liver (lane 3). Loading of RNA was estimated by ethidium bromide staining of the gel (Panel 2C).

Figure 3 outlines the strategy used for AS-30D genomic library construction and isolation of the tumor Type II hexokinase promoter. Figure 3A shows SauA I partially digested AS-30D DNA was ligated to compatible Sac I digested λ -Fix II arms to generate the genomic library. Figure 3B shows the subcloning of a 5.15 kbp promoter containing DNA fragment into vector pUC18. This 5.15 kbp DNA insert was subcloned into pGL2-Basic, a luciferase reporter vector using compatible Xba I, Nhe I sites (Figure 3C). A DNA

fragment corresponding to the coding region within the first exon was removed using Xho I, to yield a 4.3 kbp promoter-reporter gene construct. Striped bars indicate luciferase cDNA. Hatched bars indicate AS-30D DNA. Open boxes indicate the multicloning sites (mcs) and DNA of individual vectors +1, transcription start site; cos, λ- cohesive termini..

Figures 4A and 4B show the identification of λ-subclones containing the tumor Type II hexokinase promoter region. DNA from six positive λ-clones (Lanes 1-6; clones 22-1, 22-2, 22-3, 25-1, 27-1, 29-1) were digested with Xba I and separated by agarose gel electrophoresis, as shown in Figure 4A. DNA fragments similar to skeletal muscle Type II hexokinase first exon (white bars) were identified by Southern blot analysis, as shown in Figure 4B. Molecular weight markers (λ-Hind III), in kbp, are shown to the right.

Figure 5 shows the complete nucleotide sequence of the 4.3 kbp proximal promoter region, the first exon, and part of the first intron of the AS-30D tumor Type II hexokinase gene. The response element motifs are indicated below the nucleotide sequence. Sequence in italics indicate a motif with the potential to form Z-DNA structures. Direct DNA repeats larger than 10 bp are indicated by dotted lines below the sequence, and identified by Roman numerals. A TATA box motif (-30) and a CAAT box motif (-85) are highlighted and the DNA sequence of the first exon is underlined (+1 to +524). Nucleotides +525 to +790 are part of the first intron.

Figure 6 shows cDNA and deduced amino acid sequences of the Type II tumor hexokinase from rat hepatoma AS-30D. The differences between the Type II rat skeletal muscle hexokinase and the AS-30D hexokinase at the nucleotide and amino acid levels are highlighted within the sequence.

Figure 7 shows cDNA and amino acid sequences of the PCR-generated probe used for isolation of the promoter for Type II tumor hexokinase from the

AS-30D genomic library. The DNA region which corresponds to the 260 bp PCR product is underlined.

Figure 8 outlines the organization of the potential response elements for glucose, insulin, glucagon, TPA, and cAMP on the 4.3 kbp tumor Type II hexokinase promoter (+1, transcription start site; closed box, mRNA untranslated region; hatched box, coding region of the first exon; open boxes, response elements that are sensitive to glucose, insulin, glucagon, cAMP or TPA).

Figure 9 shows the effect of glucose, insulin, glucagon, TPA, and cAMP on the transcriptional activity of the tumor type II promoter. Luciferase activity was assayed 24 hours post-transfection of the promoter-reporter construct into AS-30D cells, which were maintained under hormonal, metabolite, or intracellular mediator influence as indicated in the figure. Activities are expressed as fold activation over that of a control (1mM pyruvate). Each of the samples contained 1 mM pyruvate as substrate background. All values represent the mean of six independent experiments. The individual standard deviations (± SD) for the fold activation are: glucose , 0.44; insulin, 0.25; glucagon, 0.175; glucose + insulin, 0.5; glucose + glucagon, 0.06; TPA, 0.45; and dibutyryl cAMP, 0.7.

Figure 10 shows the transcriptional activity of the tumor type II hexokinase promoter in normal or AS-30D hepatoma cells. Luciferase activity was assayed 24 hours post-transfection of the promoter-reporter construct into hepatocytes or into AS-30D hepatoma cells. Hormonal or metabolite conditions used, are indicated on the figure; activity is expressed as fold activation over that of a control (1 mM

lactate). Each sample contained 1 mM lactate as substrate background. All values represent the mean of two independent experiments.

Figure 11 outlines segments of the nucleotide sequence of the hepatoma AS-30D Type II hexokinase promoter region shown in Figure 5.

Transcription factors that bind to specific regions within the DNA sequence are indicated. The respective response elements (transcription factor binding sites) may consist of all or part of the sequence shown.

An example of how to interpret the information in Figure 11 is set forth below.

interpretation

-4369 TCTAGAGCTCGTCGCGGCCGCGGA

(+) LBP-1 SOO487

SEQUENCE POSITION = -4369 first sequence reference point

DNA SEQUENCE = TCTAGAGCTCGTCGCGGCCGCGGA

SENSE STRAND = (+)

TRANSCRIPTION FACTOR = LBP-1

REFERENCE # FOR TRANSCRIPTION FACTOR FROM TRANSCRIPTION FACTOR DATABASE (TFD) AT THE NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION = SOO487

For ease of reference, the definitions used in the specification are as indicated in Table 1.

TABLE 1. DEFINITIONS

Genomic DNA

A DNA fragment derived from cellular chromosomal DNA rather than from messenger RNA, the source for cDNA clones. Genomic and cDNA clones have different sequences due to RNA splicing. Other means of producing genomic DNA can be used.

cDNA (complementary DNA)

DNA synthesized from mRNA in test tubes using an enzyme called reverse transcriptase. The cDNA sequence is thus complementary to that of the mRNA. Complementary DNA is sometimes made with radioactive nucleotides and is used as a hybridization probe to detect specific RNA or DNA molecules. Other means of producing cDNA can be used.

Nucleotide

A monomeric unit of deoxyribonucleic acid ("DNA") or ribonucleic acid ("RNA") consisting of a sugar moiety (pentose), a phosphate, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of the pentose) and that combination of base and sugar is called a nucleoside. The base characterizes the nucleotide. The four DNA bases are adenine ("A"), guanine ("G"), cytosine ("C"), and thymine ("T"). The four RNA bases are A, G, C, and uracil ("U").

DNA Sequence

A linear array of nucleotides connected one to the other by phosphodiester bonds between the 3' and 5' carbons of adjacent pentoses.

Codon

A DNA sequence of three nucleotides (a triplet), which encodes, through its template, an amino acid, a translation start signal, or a translation termination signal. For example, each of the nucleotide triplets TTA, TTG, CTT, CTC, CTA, and CTG encode the amino acid leucine ("Leu"), TAG, TAA, TGA are translation stop signals, and ATG is a translation start signal.

Polypeptide

A linear array of amino acids connected one to the other by peptide bonds between the alpha-amino and carboxy groups of adjacent amino acids.

Transcription

The process of producing RNA from a structural gene.

Translation

The process of producing a polypeptide from mRNA.

Expression

The process undergone by a structural gene to produce a polypeptide. It is a combination of transcription and translation.

Plasmid

Nonchromosomal, double-stranded DNA sequence comprising an intact "replicon" such that the plasmid is replicated in a host cell. When the plasmid is placed within a unicellular organism, the characteristics of that organism may be changed or transformed as a result of the DNA of the plasmid. For example, a plasmid carrying the gene for neomycin restistance (neo.sup.R) transforms a cell previously sensitive to neomycin into one which is resistant to it. A host cell transformed by a plasmid or vector is called a "tranformant".

Phage or Bacteriophage

Bacterial virus which consists of nucleic acids encapsulated into a protein envelope or coat ("capsid").

Cloning Vehicle or Vector

A plasmid, phage DNA or other DNA sequence which is able to replicate in a host cell, characterized by one or a small number of endonuclease recognition or restriction sites at which such DNA sequences may be cut in a determinable fashion without attendant loss of an essential biological function of the DNA, e.g., replication, production of coat proteins, or loss of promoter or binding sites, and which contain a marker suitable for use in the identification of transformed cells, e.g., neomycin resistance or ampicillin resistance.

Molecular Cloning

The process of transferring a specific DNA fragment from one organism into another organism via a vector and stably maintaining the foreign DNA molecule in the second organism.

Recombinant DNA Molecule or Hybrid DNA

A molecule consisting of segments of non-contiguous DNA, which have been joined end-to-end.

Expression Control Sequence

A sequence of nucleotides that controls and regulates expression of genes when operatively linked to those genes.

Probe

A DNA or RNA molecule, which is used to locate a complementary RNA or DNA by hybridizing to it. Often a probe is used to identify bacterial colonies or phage plaques that contain cloned genes and to detect specific nucleic acids following separation by gel electrophoresis.

Promoter

The DNA sequence to which RNA polymerase specifically binds and at which it initiates RNA synthesis (transcription) of a specific gene.

Antisense

A DNA or RNA molecule constructed in an orientation that is complementary to the protein or mRNA coding sequence of the actual gene.

Exogenous, Foreign or Alien DNA

DNA sequences or sequence motifs not naturally present in the host genome.

Hypoxia

Reduction of oxygen supply to tissue below physiological levels despite adequate perfusion of the tissue by blood.

Glycolysis

The enzymatic conversion of glucose to the simpler compounds lactate or pyruvate. Breakdown of glucose under aerobic conditions forms pyruvate. Breakdown of glucose under anaerobic conditions forms lactic acid.

Tumor

A neoplasm or a new or abnormal tissue growth which is uncontrolled and progressive.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides the promoter and adjacent transcriptional regulatory regions for the tumor Type II hexokinase, which were isolated, sequenced, and subjected to reporter-gene analysis. Several features of the promoter are of special interest. The promoter contains response elements *inter alia* for insulin, glucose, cAMP, and TPA. In particular, insulin, glucose, glucagon, cAMP analogs, and the phorbol ester, TPA, activate the promoter. These findings suggest that signal transduction cascades involving tyrosine kinase, protein kinase A, and protein kinase C are involved in enhancing transcription of the Type II hexokinase gene in highly glycolytic tumor cells. This is in sharp contrast to the parent hepatocytes where both glycolysis and the level of Type II hexokinase isozyme are very low, and the expressed Type IV isozyme, glucokinase, although transcriptionally activated by insulin, is inhibited by glucagon and cAMP. Thus, during or subsequent to, the hepatocyte ----> hepatoma transformation, a genetic switch takes place in order to provide the hepatoma cells with a low Km (glucose) hexokinase that remains highly expressed regardless of metabolic state.

The invention relates particularly to a DNA fragment or regulatory region comprising a Type II hexokinase promoter and flanking sequences. Preferably, this promoter is derived from the genomic DNA of mammalian tumor cells, especially hepatoma cells. Optionally, the Type II hexokinase promoter is followed by all or part of the Type II hexokinase coding region naturally linked to the promoter. Particularly preferred as the first exon. Additionally, the DNA fragment may contain sequences which are required for efficient translation of mRNA. Also, mutant forms of the DNA

fragment, which retain promoter or regulatory function, are encompassed by this invention.

A regulatable promoter is a promoter where the rate of RNA polymerase binding and/or initiation is modulated by external stimuli. Stimuli may include glucose, insulin, oxygen, light, heat stress and the like. Inducible, suppressible, and repressible promoters are some types of regulatable promoters. Based upon the characterization of the disclosed promoter and its regulatory sequences, the Type II hexokinase promoter is a regulatable promoter.

For ease of reference, the abbreviations used herein are as indicated in Table 2.

TABLE 2. Abbreviations

kbp	-	kilobase pairs
IRE	-	insulin response element
ATF/CRE	-	cyclic AMP response element
Ap-1	-	activator protein-1
C/EBP	-	CCAAT-enhancer binding protein
GIRE	-	glucose response element
HNF	-	hepatocyte nuclear factor
EGF	-	epidermal growth factor
TPA	-	phorbol 12-myristate 13-acetate
pfu	-	plaque-forming units
EtBr	-	ethidium bromide
PBS	-	phosphate buffered saline
SSC	-	sodium chloride-sodium citrate solution
Glu-6-P	-	glucose-6-phosphate
cDNA	-	complementary deoxyribonucleic acid
DTT	-	dithio threitol
NIDD	-	non-insulin-dependent diabetes

Isolated DNA fragments according to the present invention are genomic DNA fragments which are no longer attached to the chromosome from which they derive. The fragments can be isolated from genomic DNA, or from genomic DNA libraries,

or synthesized according to the sequences disclosed herein. The DNA fragments can be isolated from a variety of different mammalian species using known techniques for isolating homologous sequences. The response elements contained within the fragment are highly conserved among species. The response elements are numerous in the approximately 4.3 kbp upstream from the transcription start site of tumor Type II hexokinase. All or less than all of the response elements can be used. In addition to the RNA polymerase binding site, a number of different binding sites for transcription factors are found. These can be identified in any particular sequence by comparison to the known sites (sequences) for the transcription factors. The sites in the sequence of the fragment of SEQ ID NO:1 have been found using a computer program for comparing sequences. Such programs can be applied to any particular sequence of a promoter and regulatory region of a mammalian tumor Type II hexokinase. Figure 11 lists the response elements and their locations using nucleotide coordinates. As shown below, many of these response elements observed by inspection of the sequence have been found to be functional.

The response elements of the present invention regulate the transcription of a downstream open reading frame. This has been demonstrated using reporter genes, such as luciferase. Other reporter genes, *i.e.*, genes which produce a readily observable or assayable product, can be used as is convenient. Many such genes are known in the art. Other open reading frames which can be covalently joined to the response elements of the present invention include genes encoding toxic products. These include toxins as well as enzymes which convert innocuous compounds to toxic ones. Well known examples of such toxic genes are ricin and herpes simplex virus thymidine kinase. Any such gene can be used as is appropriate for the particular application. Covalent joining of the transcription regulatory elements of the present invention to open reading frames other than for hexokinase forms recombinant constructs which are termed gene fusions.

Gene fusions according to the present invention can be used in a variety of methods. Gene fusions having a reporter gene attached to a transcriptional regulatory region can be used to screen for potential drugs. Test substances which increase

transcription of the reporter gene would be useful for treating diabetes. They could be used to increase the rate of glycolysis in the cells which would increase the absorption of glucose from the bloodstream. Conversely, test substances which decrease transcription of the reporter gene could be used to treat neoplasia, as up-regulation of glycolysis appears to be important for their unregulated growth. Test substances are contacted with the reporter gene fusions, either in an *in vitro* transcription setting or in an *in vivo* cellular situation. Transcription can be measured in the presence and absence of the test substance. Any suitable method among the many which are known in the art can be used for measuring transcription. Typical assays utilize labeled nucleotide substrates which are incorporated into transcripts and which can be quantitated. The assays can be performed in the presence of the particular stimuli which affect the transcriptional regulatory factors which bind to the transcription regulatory region of the present invention.

The gene fusions of the present invention can also be used for treating cells which overexpress hexokinase II. Many tumor cells have been found to have the high glycolytic rate phenotype. Thus a treatment method can take advantage of this by using a transcription regulatory region which is very active in the particular cells which one desires to treat, namely that of tumor type II hexokinase. According to one aspect of the invention, humans and mammals are treated when they have been determined to have tumor cells with an increased rate of glucose utilization over normal cells and/or the capacity to sustain high rates of glycolysis under aerobic (solution or physiological fluid saturated with dissolved oxygen at room temperature (25 degrees Celsius)), hypoxic or under conditions known as hypoxia (low oxygen levels and/or a solution or physiological condition having less saturated conditions as compared to aerobic conditions but not reaching anaerobic conditions), or anerobic conditions (solution or physiological fluid having extremely low [near zero] oxygen levels but not hypoxic levels). Another benefit of using the response elements of the present invention is that they appear to be active in tumor cells but not in normal cells. Thus they provide a means of selectively expressing genes in tumor cells. While not wishing to be bound by any particular theory, one cause of the selective expression may be the

presence of mutant forms of p53 in tumor cells. According to the method of the invention, a gene fusion, which comprises a toxic protein regulated by a transcription regulatory region of the present invention, is administered to cells which overexpress hexokinase II. The toxic gene is thus expressed at a high rate and selectively in the tumor cells. Suitable vectors for administering such fusion genes are known in the art and include without limitation, adenoviruses, retroviruses, herpes viruses, and minichromosomes. Delivery methods as are known in the art can be used, including liposomes, infection, inhalation, etc.

Nucleic acid probes of the present invention comprise at least 15 contiguous nucleotides selected from the sequence of SEQ ID NO:1. Preferably they are selected from the regulatory region comprising nucleotides -4369 to -1. More preferably they are selected from nucleotides number -4369 to -1158. However, other sequences including those of the first exon may be included. Nucleic acid probes may or may not be labeled with a detectable label, including but not limited to a radiolabel, a fluor, and an enzyme.

According to another aspect of the invention, the amplification of the hexokinase type II gene in tumor cells can be used as a diagnostic marker of neoplasia and as a prognostic marker of an aggressive type of tumor. Amplification is any gene copy number greater than 2, although copy numbers of greater than 4, 6, 8, and 10 are possible. Determination of copy number can be by any means known in the art, including fluorescence in situ hybridization, quantitative polymerase chain reactions, quantitative Southern blotting. A suitable control sample can be used from a somatic tissue which is observed to be morphologically normal.

Another diagnostic method of the present invention stems from the observation that the type II hexokinase gene is methylated and expression-silent in normal cells but unmethylated and heavily expressed in tumor cells. Thus by determining whether cells in a tissue sample contain an unmethylated or methylated hexokinase II gene one can ascertain neoplasia and/or aggressive tumor behavior. Any test known in the art for determining DNA methylation can be used. For example, restriction endonucleases like *DpnI* which only cleave sequences containing methylated adenine or cytosine can

be used to distinguish between methylated and unmethylated sequences. Methylation foot printing of the gene can also be used to determine the methylation status of the gene.

Vectors are provided which employ the transcription regulatory fragment of the present invention for expression of a desired protein in a mammalian cell. Because of the multiple means of regulating the tumor type II hexokinase, this system provides a very attractive tool for manipulation of expression levels and conditions. Vectors according to the invention include without limitation viral and plasmid vectors as are known in the art.

According to another aspect of the invention, humans and mammals are treated when they have been determined to have cells with a decreased level or rate of glucose utilization over normal cells and/or an incapacity to sustain high rates of glycolysis under aerobic conditions, such as in non-insulin-dependent diabetes. Even in cases where glucase utilization rates are normal treatment to increase the rate can benefit the patient by removing excess glucose from the blood. Glycolysis in cells is increased by transfecting them with an unmethylated copy of the tumor type II hexokinase gene. Unmethylated copies can be obtained, *inter alia,* by passaging the DNA though a non-methylating host. The gene comprises both the regulatory region and the coding region of the hexokinase. By increasing the copy number of the gene, the rate of glycolysis increases, thereby increasing the amount of glucose which is absorbed by the cells from the bloodstream. Particularly suitable cells for such treatment include muscle, liver, and adipose cells.

The disclosed DNA sequence encoding the tumor Type II hexokinase promoter may be synthesized chemically or isolated by one of several approaches well known to one skilled in the art. For example, the complete sequence may be assembled from overlapping oligonucleotides prepared by standard methods and assembled into a complete sequence (Edge, (1981), Nature, 292:756; Nambair, et al., (1984), Science, 223:1299; Jay, et al., (1984), J. Biol. Chem., 259:6311). Isolation methods may include nucleic acid hybridization using appropriate single stranded or double stranded DNA or oligonucleotide probes. Such probes can be constructed synthetically, based

upon the DNA or amino acid sequences disclosed herein, or isolated from genomic or cDNA clones also described herein.

The preparation of oligonucleotides and DNA libraries, as well as screening by nucleic acid hybridization, are well known in the art (Sambrook, supra). Standard hybridizing conditions and procedures are known in the art (Southern, E., (1975), J. Mol. Biol., 98:503); Sambrook, supra). If the nucleic acid is mRNA, it is contacted with a labeled Type II hexokinase gene probe complementary to the RNA under standard hybridizing procedures. The probe may be DNA, cDNA, or RNA depending upon the nucleic acid extracted and the method of hybridization chosen. The probe may be part of the sequence of the Type II hexokinase gene, including all coding and non-coding regions, a sequence including only the coding or non-coding regions, or any fragment(s) thereof. Preferably, the probe is tumor Type II hexokinase cDNA. The nucleic acid also may be amplified using PCR or RT-PCR prior to contact with the probe.

The construction of a DNA library is well known to the skilled artisan. The library can, for example, consist of a genomic library from a human source. Preferably, the DNA libraries are constructed of chromosomal DNA. The genomic DNA or cDNA is cloned into a vector suitable for construction of a library.

Once the library is constructed, oligonucleotides or other DNA or RNA molecules may be used to probe the library to identify the segment carrying the sequence encoding the Type II hexokinase promoter in the case of a gene library. Oligonucleotides can be designed and produced for use as hybridization probes to locate the sequences encoding the promoter. In general, the probes are synthesized chemically, preferably based upon known nucleic sequences, such as the 260 bp probe as shown in Figure 7. Ultimately, the isolated segments of DNA are ligated together so the correct sequence is constructed.

In designing probes, the nucleotide sequences can be selected as to correspond to codons encoding the amino acid sequence. Since the genetic code is redundant, degenerate probes include several oligonucleotides to cover all, or a reasonable number, of the possible nucleotide sequences, which encode a particular amino acid

sequence. Thus, it is generally preferred, in selecting a region of the sequence upon which to base degenerate probes, that the region not contain amino acids whose codons are highly degenerate (Lathe, (1985), J. Mol. Biol., 183:1-12; Sambrook, supra).

The assembled sequence can be cloned into any suitable vector or replicon and maintained there in a composition which is substantially free of vectors that do not contain the assembled sequence. This provides a reservoir of the assembled sequence, and segments or the entire sequence can be isolated from the reservoir by excision with restriction enzymes or by polymerase chain reaction (PCR) amplification. The polymerase chain reaction is performed by methods and conditions disclosed in U.S. Patent Nos. 4,683,202 and 4,683,195, Sambrook, surpa, and in Perkin Elmer Cetus PCR kit protocols. The DNA polymerase, deoxyribonucleotide triphosphates (dNTPS) (e.g, dATP, dCTP, dTTP, and dGTP), and amplification buffer (e.g., glycerol, tris-hydrochloric acid, potassium chloride, Tween 20, and magnesium chloride) are commercially available (Perkin Elmer Cetus). The amplification process may be performed for as many cycles as desired. Numerous cloning vectors are known to those skilled in the art, and the selection of an appropriate cloning vector is a matter of choice (Sambrook, supra). The Type II hexokinase promoter and/or gene sense and antisense primers for use in PCR may be selected from primers described herein or others synthesized from the Type II hexokinase promoter and/or gene. The primers may be produced using a commercially available oligonucleotide synthesizer, such as Applied Biosystems Model 392 DNA/RNA synthesizer. Either the sense or antisense primer may be labeled with a detectable marker by known procedures such as phosphorylation with bacteriophage T4 polynucleotide kinase (Sambrook, supra). Suitable markers include, but are not limited to, fluorescence, enzyme, or radiolabels such as ^{32}P and biotin.

An expression vehicle may be any vector which is capable of transfecting mammalian cells and expressing a desired gene. The gene may encode a therapeutic agent for treating cells in need of such therapy either in vitro or in vivo. Suitable expression vehicles which may be employed include, but are not limited to, eukaryotic vectors, prokaryotic vectors, and viral vectors, such as adenovirus vectors, adeno-

associated viral vectors, retroviral vectors (e.g., Moloney Murine Luekemia Virus, vectors derived from retrovriuses such as Rous Sarcoma Virus), herpes virus vectors, DNA-protein complexes, and receptor-mediated vectors. Any such vector may be contained within a liposome. Preferably, the vector of choice includes the tumor Type II hexokinase promoter. Other suitable promoters, however, that may be employed include, but are not limited to, the retroviral LTR, the SV40 promoter, and the human cytomegalovirus (CMV) promoter (Miller, et al., (1989) *Biotechniques, 7*:980-990).

The construction of vectors containing desired DNA segments linked to appropriate DNA sequences is accomplished by techniques similar to those used to construct the segments. These vectors may be constructed to contain additional DNA segments, such as bacterial origins of replication to make shuttle vectors (for shuttling between prokaryotic hosts and mammalian hosts).

Procedures for construction of mutant sequences are well known in the art. A DNA sequence encoding a mutant form of Type II hexokinase promoter, for example, can be synthesized chemically or prepared from the wild-type sequence by several techniques, *e.g.*, primer extension, linker insertion, and PCR (Sambrook, supra). Mutants can be prepared which have deletions, substitutions, and insertions relative to the wild-type sequence. Confirmation of specific mutant sequences can be conducted by sequence analysis and/or assays described herein.

Nucleic acids can be extracted from desired cells by known techniques. The nucleic acids may comprise DNA or RNA, preferably, genomic DNA or mRNA. For example, specific cells can be lysed using proteinase K in the presence of detergents, such as sodium dodecyl sulfate (SDS), NP40, or Tween 20. If the nucleic acid is genomic DNA, it is then extracted using known techniques, such as phenol/chloroform extraction, or other procedures (U.S. Patent Nos. 4,900,677 and 5,047,345). Alternatively, DNA can be isolated using one of the commercially available kits, such as Oncor Genomic DNA isolation kit. RNA can be extracted using various known procedures, such as guanidinium thiocyanate followed by centrifugation in cesium chloride (Sambrook, supra). Specifically, genomic DNA can be isolated from a cancer cell line known as hepatoma AS-30D. A method for culturing cells is described in

Example 1, this and other methods are well known in the art. The DNA can be restricted into smaller-sized fragments of 10 to 20 kbp using the restriction enzyme Sau3 AI. To generate a random population of such fragments for laboratory manipulations, the DNA fragments can be placed in viruses. For this purpose, the DNA fragments can be ligated to viral DNA isolated from the bacteriophage λ-Fix II (Stratagene Cloning Systems, La Jolla, CA). A sub-population of the genomic library cana be screened using a DNA probe synthesized using the Type II hexokinase cDNA of rat skeletal muscle (Thelen and Wilson (1991), *Arch. Biochem Biophys.*, 286:645-651), using plaque hybridization and detection (Sambrook, supra). Viral particles that are identified by this technique, as containing the DNA fragments that harbor DNA similar to the DNA probe, can be isolated. DNA within these viral particlesca be released by digestion with Proteinase K and purified by selective precipitation. To isolate and identify the DNA element described in this invention, the DNA can be digested with restriction enzyme Xba I, and the digested DNA fragments can be separated using agarose gel electrophoresis and detected by Southern hybridization, again using a PCR amplified 260 bp probe corresponding to part of the rat skeletal muscle hexokinase Type II cDNA. DNA sequencing of the identified fragments can serve to locate the promoter region, the hexokinase coding regions (exons) and, additionally, the restriction site(s) which may be useful in further processing, for example, for cutting off DNA sequences which are not necessary for promoter function. Depending on the choice of restriction enzyme, the DNA fragments containing the Type II hexokinase promoter and/or transcriptional regulatory regions may also include at the 3' and 5' termini original flanking DNA sequences which do not affect the promoter function and may be used as connecting sequences in the subsequent cloning procedures. If desired, these sequences can be ligated to chemically synthesized DNA linkers, which preferably include the recognition sequence of an appropriate restriction enzyme. This allows a convenient connection of the Type II hexokinase promoter and/or transcriptional regulatory region with foreign polypeptide coding regions. It is also possible to isolate and/or construct a DNA fragment which contains the Type II hexokinase promoter and part or all of the adjacent signal sequence

from the Type II hexokinase protein coding region. When ligated to an appropriately cut foreign polypeptide coding region, the resulting hybrid DNA will be expressed in appropriate expression systems to yield desired polypeptides. The polypeptide coding region controlled by the promoter may be derived from genomic DNA or from cDNA prepared via the mRNA route or may be synthesized chemically. The isolated DNA element of 5150 base pairs was placed in a plasmid vector pUC18, for further laboratory manipulations. A vector is a replicon, such as a plasmid, phage or cosmid, into which another DNA segment may be attached so as to bring about the replication of the attached segment. For example, useful vectors may comprise segments of chromosomal DNA, non-chromosomal DNA, such as various known derivatives of SV40 and bacterial plasmids (e.g, plasmids from E. coli including pBR322, pBluescript (Stratagene), pGEM (Promega), pUC118, pUC119, pUC18, and pUC19), or synthetic DNA sequences, phage DNAs (M13) including derivatives of phage (e.g., NM 989), vectors useful in yeasts, vectors useful in eukaryotic cells, such as vectors useful in animal cells, (e.g., those containing SV40, adenovirus and retrovirus-derived DNA sequences), and vectors derived from combinations of plasmids and phage DNA (such as plasmids which have been modified to employ phage DNA), or other derivatives thereof.

The nucleotide structure of a 5150 bp DNA element comprising the tumor type II hexokinase promoter was determined by the Sanger method of dideoxy-mediated chain termination. Alternatively, the nucleic acid may be sequenced using the Maxam-Gilbert chemical degradation of DNA method (Sambrook, supra), or other procedures known to those skilled in the art. The nucleotide sequence of the promoter region includes the disclosed sequence as disclosed in Figure 5 and conservative variations thereof. In a preferred embodiment, the nuleotides are -4369 to -1, or regulatory fragments contained therein. Regulatory fragments are defined as critical fragments which regulate expression of Type II hexokinase. Figure 6 shows the deduced amino acid sequence for the As-30D hepatoma Type II hexokinase.

The overproduction of hexokinase in cancer cells correlates with markedly elevated mRNA levels. (Johansson, et al., (1985), *Biochem. Biophys. Res. Commun.*,

133:608-613; Paggi, et al., (1991), *Biochem. Biophys. Res. Commun.*, 178:648-655; Shinohara, et al., *FEBS Lett.*, 291:55-57). Northern blot analysis was carried out on total RNA to determine the expression levels of Type I and Type II mRNA hexokinase in AS-30D hepatoma cells relative to their expression in normal rat liver (Figure 2). The level of Type II hexokinase mRNA may be determined by methods well known in the art, such as Northern blotting, dot and slot hybridization, S1 nuclease assay, or ribonuclease protection assays (Sambrook, supra).

In Northern blotting, the RNA is separated by known techniques (Chirgwin, et al., (1979), Biochemistry, 18:5294) and transferred to an activated cellulose, nitrocellulose, or nylon membrane. The mRNA is then hybridized with a radiolabeled DNA or RNA probe followed by autoradiography. The probe may be the full length Type II hexokinase gene or fragments thereof. Preferably, the probe is the full length Type II hexokinase cDNA. In dot and slot hybridization, the RNA is hybridized to an excess of a radiolabeled Type II hexokinase DNA or RNA probe (Kafatos, et al., (1979), Nucleic Acids Res., 7:1541; Thomas, R, (1980), Proc. Natl. Acad. Sci., 77:5201; White, et al., (1982), J. Biol. Chem., 257:8569). The amount of the Type II hexokinase mRNA can then be determined by densitometric tracing of the audioradiograph and comparison to the amount of normal Type II hexokinase mRNA. In S1 nuclease assay or ribonuclease protection assay, RNA is hybridized with labeled DNA or RNA probes derived from genomic DNA (Berk, et al., (1977), Cell, 12:721; Casey, et al, (1977), Nucleic Acids Res.

Transgenic organisms, such as transgenic mammals, transgenic mice, transgenic fish, etc., may be formed by introducing the nucleic acid molecule of the present invention into their genome, i.e., a Type II hexokinase gene or any other gene regulated by the disclosed promoter or a mutated Type II hexokinase gene or promoter, or variations thereof. Preferably, the transgenic animal is a mouse. Methods for producing transgenic organisms containing a recombinant nucleic acid molecule are well known in the art (Alberts, et al., (1989), Molecular Biology of the Cell, 2d., Garland Publishing Inc., New York, pgs. 267-269; Gasser, et al., (1989), Science, 244:1293-1299; European Patent Application No. 0257472, filed Aug. 13, 1987 by De

La Pena, et al.; PCT Pub. No. WO 88/02405, filed Oct 1, 1987, by Trulson, et al.; PCT Pub. No. WO 87/00551, filed Jul. 16, 1986, by Verma; Wagner, et al., U.S. Patent No. 4,873,191; Rogers, et al., (1987), Meth. in Enzymol., 153:253-277; Cocking, et al., (1987), Science, 236:1259-1262; Burton, et al., U.S. Patent No. 5,416,017).

Methods and compositions for regulating the expression of foreign or alien DNA in a host, such as a mammalian host, are described below. A powerful promoter, such as the tumor Type II hexokinase promoter, is also useful for high levels of protein production. In addition, the promoter, which is regulated by a number of transcription factors, may be manipulated to enhance and/or modify protein production.

The compositions of the present invention, i.e., specific sense or antisense sequences, preferably, critical regulatory or response elements or sequences, may be made into pharmaceutical compositions with appropriate pharmaceutically acceptable carriers or diluents. If appropriate, pharmaceutical compositions may be formulated into preparations including, but not limited to, solid, semi-solid, liquid, or gaseous forms, such as tablets, capsules, powders, granules, ointments, solutions, suppositories, injections, inhalants, and aerosols, in the usual ways for their respective route of administration. Methods known in the art can be utilized to prevent release or absorption of the composition until it reaches the target organ or to ensure time-release of the composition. A pharmaceutically-acceptable form should be employed which does not inactivate the compositions of the present invention. In pharmaceutical dosage forms, the compositions may be used alone or in appropriate association, as well as in combination with, other pharmaceutically-active compounds. For example, in applying the method of the present invention for delivery of a nucleic comprising a Type II hexokinase promoter and/or gene-related elements, such delivery may be employed in conjunction with other means of treatment of cancer or diabetes, for example.

Accordingly, the pharmaceutical compositions of the present invention can be delivered via various routes and to various sites in an animal body to achieve a particular effect. Local or systemic delivery can be accomplished by administration

comprising application or instillation of the formulation into body cavities, inhalation, or insufflation of an aerosol, or by parenteral introduction, comprising intramuscular, intravenous, peritoneal, subcutaneous intradermal, as well as topical administration.

The composition of the present invention can be provided in unit dosage form, wherein each dosage unit, e.g., a teaspoon, tablet, solution, or suppository, contains a predetermined amount of the composition, alone or in appropriate combination with other pharmaceutically-active agents. The term "unit dosage form" refers to physically discrete units suitable as unitary dosages for human and animal subjects, each unit containing a predetermined quantity of the composition of the present invention, alone or in combination with other active agents, calculated in an amount sufficient to produce the desired effect, in association with a pharmaceutically-acceptable diluent, carrier (e.g., liquid carrier such as a saline solution, a buffer solution, or other physiological aqueous solution), or vehicle, where appropriate. The specifications for the novel unit dosage forms of the present invention depend on the particular effect to be achieved and the particular pharmacodynamics associated with the pharmaceutical compositon in the particular host.

Additionally, the present invention specifically provides a method of transferring nucleic acids to a host, which comprises administering the composition of the present invention using any of the aforementioned routes of administration or alternative routes known to those skilled in the art and appropriate for the particular application. The "effective amount" of the composition is such as to produce the desired effect in a host which can be monitored using several end-points known to those skilled in the art. For example, one desired effect might comprise effective nucleic acid transfer to a host cell. Such transfer could be monitored in terms of a therapeutic effect, e.g., alleviation of some symptom associated with the disease being treated, or further evidence of the transferred gene or expression of the gene within the host, e.g, using PCR, Northern or Southern hybridization techniques, or transcription assays to detect the nucleic acid in host cells, or using immunoblot analysis, antibody-mediated detection, or particularized assays, as described in the examples, to detect protein or polypeptide encoded by the transferred nucleic acid, or impacted level or

function due to such transfer. These methods described are by no means all-inclusive, and further methods to suit the specific application will be apparent to the ordinary skilled artisan.

Furthermore, the amounts of each active agent included in the compositions employed in the examples described herein, *i.e.*, add range, provide general guidance of the range of each component to be utilized by the practitioner upon optimizing the method of the present invention for practice either *in vitro* or *in vivo*. Moreover, such ranges by no means preclude use of a higher or lower amount of a component, as might be warranted in a particular application. For example, the actual dose and schedule may vary depending on whether the compositions are administered in combination with other pharmaceutical compositions, or depending on interindividual differences in pharmacokinetics, drug disposition, and metabolism. Similarly, amounts may vary *in vitro* applications depending on the particular cell line utilized, e.g., the ability of the plasmid employed for nucleic acid transfer to replicate in that cell line. Furthermore, the amount of nucleic acid to be added per cell or treatment will likely vary with the length and stability of the nucleic acid, as well as the nature of the sequence, and is particularly a parameter which needs to be determined empirically, and may be altered due to factors not inherent to the method of the present invention, *e.g.*, the cost associated with synthesis. One skilled in the art can easily make any necessary adjustments in accordance with the necessities of the particular situation.

The following examples are to aid in the understanding of the invention, and should not be construed in any way as limiting its scope.

EXAMPLE 1: EXPRESSION OF HEXOKINASE mRNA IN THE HIGHLY GLYCOLYTIC AS-30D HEPATOMA CELL LINE

Northern blot analysis was carried out on total RNA to determine the expression levels of Type I and Type II hexokinase in AS-30D hepatoma cells relative to their expression in normal rat liver (Figure 2). The level of Type II hexokinase mRNA may be determined by methods well known in the art, such as Northern blotting, dot and

slot hybridization, S1 nuclease assay, or ribonuclease protection assays (Sambrook, supra). In northern blotting, the RNA is separated by known techniques (Chirgwin, et al., (1979), Biochemistry, 18:5294) and transferred to an activated cellulose, nitrocellulose, or nylon membrane. The mRNA is then hybridized with a radiolabeled DNA or RNA probe followed by autoradiography. The probe may be the full length Type II hexokinase promoter and/or gene or fragment thereof. Preferably, the probe is the full length Type II hexokinase cDNA. In dot and slot hybridization, the RNA is hybridized to an excess of a radiolabeled Type II hexokinase DNA or RNA probe (Kafatos, et al., (1979), Nucleic Acids Res., 7:1541; Thomas, R, (1980), Proc. Natl. Acad. Sci., 77:5201; White, et al., (1982), J. Biol. Chem., 257:8569). The amount of the Type II hexokinase mRNA can then be determined by densitometric tracing of the audioradiograph and compared to the amount of normal Type II hexokinase mRNA. In S1 nuclease assay or ribonuclease protection assay, the RNA is hybridized with labeled DNA or RNA probes derived from genomic DNA (Berk, et al., (1977), Cell, 12:721; Casey, et al, (1977), Nucleic Acids Res., 4:1539). The products of hybridization are then digested with nuclease S1 or RNAase under conditions favoring digestion of single stranded nucleic acids. The amount of the Type II hexokinase fragments can then be measured by electrophoresis and compared to the size of normal Type II hexokinase mRNA fragments. In another embodiment, mutated Type II hexokinase promoter and/or gene may be determined using single stranded conformation polymorphism analysis (Orita, et al., (1989), Proc. Natl. Acad. Sci., 86:2766-2770) and PCR.

Specifically, isolation of total RNA from hepatocytes, AS-30D cells, and skeletal muscle was performed by phenol-chloroform extraction (Chromczynski, et al., (1987), *Anal. Biochem.*, 162:156-159). Total RNA (20 μg) was size fractionated on a 1.2% agarose formaldehyde gel (Sambrook, et al., (1989), In: *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), transferred to a nylon membrane in alkaline conditions by downward capillary blotting (Chromczynski, P., (1992), Anal. Biochem., 201:134-139), and then UV-crosslinked. The gel was stained with ethidium bromide to verify equal RNA sample

loading and transfer. An α-[^{33}]PdATP labeled probe corresponding to the full length Type II cDNA hexokinase from skeletal muscle (Thelen and Wilson, (1991), Arch. Biochem. Biophys., 286: 645-651) was used to detect the Type Iike mRNA in the blot. For Type I hexokinase, an EcoRI/BamH I fragment of the c37 mouse hepatoma Type I hexokinase cDNA (Arora, et al., (1990), *J. Biol. Chem.*, 265:6481-6488) was used as the probe (Figure 7), and for Type II hexokinase a full length cDNA rat skeletal muscle Type II hexokinase (Thelen, supra) was used. Both probes showed specific hybridization bands with rat brain and skeletal muscle RNA (Figures 1A and 1B) used as positive controls for Type I and Type II hexokinase, respectively. Hybridization bands were visualized by autoradiography on day 5 (-70 degrees Celsius). Both isozymes could be detected easily in AS-30D cells. Type II hexokinase, however, showed a much stronger hybridization signal as compared to the Type I isozyme. Neither hexokinase transcript could be detected in the normal rat liver.

AS-30D hepatoma cells were propagated in female Sprague-Dawley rats (Nakashima, supra; Arora, supra). The hepatoma cells, in ascitic form, were harvested 6 to 7 days post-transplantation. Hepatocytes were isolated from female Sprague-Dawley rats (200-250 g) by collagenase perfusion (Freshney, R., (1987), In: Culture of Animal Cells: A Manual of Basic Technique, 2nd Ed., Wiley-Liss, New York, pgs. 264-265). After perfusing the liver, the hepatocytes were resuspended in 20 ml of RPMI-1640 medium. An equal volume of 90% (v/v) Percoll solution (17 mM NaCl, 5.4 mM KCl, 81.3 mM MgSO4, 1 mM phosphate buffer, pH 7.4) (Berry et al., (1991), Lab. Tech. Biochem. Mol. Biol., 21:24-25, 56-57) was added and mixed. The viable hepatocytes were separated by centrifugation (50 x g, 5 min.) and washed once in RPMI-1640 medium.

EXAMPLE 2: ISOLATION OF THE TUMOR TYPE II HEXOKINASE PROMOTER

To identify the cis-transcriptional control elements that regulate the expression of the tumor Type II hexokinase gene, a 5.1 kbp genomic clone containing the

proximal promoter region and the first exon of the tumor Type II hexokinase was isolated and mapped. Genomic DNA was isolated from AS-30D hepatoma cells (Sambrook, supra). Partial digestion of the genomic DNA with Say-3AI to generate 10 to 20 kbp DNA fragments and partial fill-in of the Sau-3AI ends with Klenow fragment to create ends which are incompatible with each other, but are complimentary to Xho I partially filled ends (Promega Protocols and Applications Guide, (1991), 2nd Ed., Promega Corp., Madison, WI). Isolation of λ-Fix II phage DNA by a liquid lysate method and modification of λ-Fix II phage DNA to generate Xho I half-site arms, and ligation of the half-site λ-arms to the partially filled-in AS-30D genomic DNA. The ligated DNA was packaged in vivo using Gigapack II Gold packaging extract (Stratagene). The recombinants were screened on duplicate nitrocellulose membranes (132 mm) (Schleicher and Schuell) at a density of 5 x 10^4 pfu/plate. For the isolation, approximately 5 x 10^5 plaques were screened from an unamplified AS-30D genomic library packaged at an efficiciency of 6.6 x 10^6 pfu/ug. Six plaques, which hybridized to a 260 bp PCR, as shown in Figure 7, amplified probe corresponding to a part of the Type II skeletal muscle hexokinase first exon, were isolated as shown in Figure 2A. The recombinant phage, denoted 22-1, 22-2, 22-3, 25-1, 27-1, and 29-1, contained genomic DNA inserts in the size range of 10 kbp to 20 kbp. Recombinant λ-DNA prepared from each isolate was completely digested with Xba I. DNA fragments containing sequences corresponding to the first exon of skeletal muscle Type II hexokinase were identified by agarose gel electrophoresis, as shown in Figure 3A, followed by Southern hybridization, Figure 4B, using the 260 bp PCR product as probe. The 260 bp PCR generated DNA fragment corresponding to the positions -197 to +63 of Type II hexokinase (translation start point referenced as +1) (Thelen, supra) was [α-^{32}P]dATP radiolabeled by nick translation and used as the probe to screen the library. Specific Xba I digested DNA fragments, in the size range 0.75 kbp (page 22-10, 5 kbp (phage 22-2 and 29-1), and 9 kbp (phage 22-3, 25-1, and 27-1), were isolated from individual λ-clones and subcloned into the plasmid vector pUC18. A pUC18 subclone containing the approximately 5 kbp DNA insert from the λ-clone 29-1, was sequenced at the termini using pUC 18 universal primers to test for

the presence of DNA similar to the Type II hexokinase first exon, as described below in Example 3, and selected for further characterization.

EXAMPLE 3: SEQUENCING AND ANALYSES OF THE TUMOR TYPE II HEXOKINASE PROMOTER

The 5 kbp subclone, 29-1/Xba I, was sequenced in both directions as shown in Figure 3B to yield a full length DNA sequence of 5150 bp (Figure 5) containing the promoter, the first exon, and part of the first intron. Analysis of the putative first exon and comparison with the published sequences for the first exon of the Type II hexokinase from adipose tissue (Printz, et al., (1993), *J. Biol. Chem.*, 268:5209-5219) and skeletal muscle (Thelen, supra) indicated that the corresponding regions within the AS-30D tumor Type II hexokinase are very similar. The 5.1 kbp subclone contained a 257 bp segment of the first intron, a 63 bp coding region of the first exon, a 461 bp untranslated region of the first exon, and a 4369 bp proximal promoter region (Figure 5).

The promoter sequence was analyzed for response elements using available databases (Fasisst, et al., (1992), *Nucleic Acids Res.*, 20:3-26; Locker, et al., (1993), In: *Gene Transcription: A Practical Approach* (Hames, B.D., and Higgins, S.J. eds.), Oxford U. Press, New York, NY. Numerous response elements found within the promoter by computer analysis are indicated in Figure 5 below the DNA sequence and in Figure 11. Response elements that are sensitive to two of the main signal transduction cascades, the protein kinase A and protein kinase C pathways, and to insulin, glucagon, and glucose are indicated in Figure 8. Many DNA direct repeats, ranging from 7 bp to 36 bp, were found within the promoter. Those which are longer than 10 bp are indicated. Another interesting motif found within the distal part of the promoter was a 31 bp 'T-G' pyrimidine-purine repeat (Figure 5).

In summary, numerous response elements or regions of potential relevance to the transcriptional regulation of the tumor Type II hexokinase gene were found, including those for well established regulators of carbohydrate metabolism.

EXAMPLE 4: FUNCTIONAL ACTIVITY OF THE TUMOR TYPE II HEXOKINASE PROMOTER IN THE PRESENCE OF KNOWN REGULATORS OF CARBOHYDRATE METABOLISM

The functional activity of the tumor Type II hexokinase promoter in the presence of potential modulators of greatest interest was examined. Using a reporter gene construct consisting of the tumor Type II hexokinase promoter and the luciferase gene, the relative activity of the promoter in driving transcription in the presence of glucose, insulin, glucagon, dibutyryl cAMP, and TPA was tested.

The 4.3 kbp promoter was placed in the pGL2-Basic reporter vector, shown in Figure 4C), which is designed to test a promoter's activity by using a luciferase as a reporter gene. The promoterless luciferase plasmid vector, pGL2-Basic, was used for all promoter studies. An additional promoter-probe vector, such as the promoterless chloramphenicol acetyl transferase (CAT), could be used to test the promoter activity of the Type II hexokinase promoter. An SV-40 promoter-β-galactosidase reporter vector (pSV-β -galactosidase) was used as an internal control for evaluating the efficiciency of transfection in each experiment. An SV-40 promoter-luciferase reporter vector (pGL2-Control) was used to evaluate the transcription strength of the tumor Type II hexokinase promoter. The xba I digested DNA fragment, which contained the proximal promoter region and the first exon of the AS-30D tumor Type II hexokinase gene, identified by DNA sequencing, was inserted into the compatible Nhe I site of the luciferase reporter plasmid pGL2-Basic, upstream of the luciferase cDNA. This construct was sequenced at the sites of ligation using synthetic oligonucleotides to verify orientation and accuracy of ligation. A part of the first exon, including the coding region of tumor Type II hexokinase, was excised from the reporter construct by Xho I digestion followed by religation. The tumor Type II hexokinase promoter-reporter construct (10 μg) was transfected with 2.5 μg of the pSV-β-galactosidase vector into AS-30D hepatoma cells using 25 x 10^6 cells in 0.5 ml per transfection. Hepatocytes were transfected with DNA using 20 x 10^6 in 0.5 ml per transfection. Briefly, the cells and plasmid DNA were incubated on ice for 10 minutes and electroporated at 200 volts, 800 μF. After 10 additional minutes on ice, the cells were

plated into 10 ml of RPMI-160 glucose-deficient media (pH 7.4) supplemented with an antibiotic-antimycotic mixture, 25 mM Hepes, and 1 mM sodium pyruvate or 1 mM sodium lactate. Based on the transfection study, individual cell samples were further supplemented with 25 mM glucose, 100 mM bovine insulin, 10 μM glucagon, 100 μM dibutyryl cAMP, 100 nM TPA, or combinations thereof. The transfected cells were incubated at 37 degrees in 5 percent carbon dioxide. Cell extracts were prepared 24 hours post-transfection using cold lysis buffer (0.625% Triton X-100, 01. M potassium phosphate, and 1 mM DTT, pH 7.8) (Showe et al., (1992), Nucleic Acids Res., 20:3153-3157). Luciferase activity in the cell lysates was measured as relative light units (RLU) using standard methods (de Wet, et al., (1987), Mol. Cell Biol., 7:725-737; Turner TD020e Luminometer (Turner Designs); Promega Luciferase Assay System Kit).

AS-30D cells were chosen for the transient gene expression study, to ensure the presence of signal transduction cascades and cell-surface receptors characteristic of the parental tumor line. Transient expression of luciferase derived from the promoter-reporter construct was determined after transfection of AS-30D cells, by assaying luciferase activity 24 hours post-infection. Luciferase activity was normalized to the β-galactosidase activity derived from the co-transfected internal control plasmid pSV-β-galactosidase to correct for differences in transfection efficiency. The fold activation of the promoter was based on the activity observed when the transfected cells were maintained in 1 mM pyruvate containing RPMI-1640 medium (control). Under these "background" conditions, and in the presence of 10% serum, the tumor Type II hexokinase promoter supported significant levels of transcription comparable to that of an SV 40 promoter (data not shown).

The relative activity of the tumor Type II hexokinase promoter in driving transcription in the presence of glucose, insulin, glucagon was then tested. Preliminary studies using a 1 mM lactate or 1mM pyruvate substrate background indicated that glucose, insulin, and glucagon were capable of directing expression, where the levels of expression observed for each component were similar regardless of the lactate or pyruvate substrate background. Detailed studies using six independent experiments,

carried out in a substrate background of 1 mM pyruvate, as shown in Figure 9. The highest activation of the promoter was observed in the presence of both insulin (100 nM) and glucose (25 mM), with a 4.3 fold increase in activity. Separately, glucose and insulin gave activation levels of 3.4 and 2.4 fold, respectively. Glucagon alone caused a moderate but reproducible activation (1.3 fold) for promoter activity, which increased to 2.4 fold in the presence of glucose. Insulin and glucagon together activated the promoter by 2.8 fold which was 0.4 fold above the transcription enhancement observed in the presence of insulin alone.

The effect of analogs that activate two of the major signal transduction pathways, namely the protein kinase A and protein kinase C signaling cascades, on the tumor Type II hexokinase promoter was tested using dibutyryl cAMP (100 μM), and analog of cAMP, and TPA (100 nM), an analog of diacylglycerol, respectively. These analogs increased promoter activity by 2.1 fold and 3.3 fold, respectively. These findings emphasize the promiscuity of the tumor Type II hexokinase promoter in its activation response to a wide variety of known modulators of carbohydrate metabolism. Thus, in hepatocytes, where insulin, glucose, and glucagon are all known to regulate the expression of Type IV hexokinase, these same agents produced little or no effect on the activity of the transfected Type II tumor hexokinase promoter. These results implicate the presence of one or more unique transcription factors essential for the activation of the Type II hexokinase promoter, and therefore, the overexpression of the Type II enzyme in hepatoma cells. The significance of these findings emphasize that normal versus tumor cell differences in the regulation of hexokinase genes involved in glucose catabolism, and indicate that transcription of the Type II tumor gene may occur independent of metabolic state. Thereby providing the cancer cell with a selective advantage over its cell of origin in the production of the key metabolic precursor Glu-6-P.

EXAMPLE 5: RELATIVE ACTIVITY OF TUMOR TYPE II HEXOKINASE PROMOTER IN HEPATOCYTES AND IN AS-30D HEPATOMA CELLS

The reporter vector was transfected into hepatocytes in order to test whether the tumor Type II hexokinase promoter was capable of driving transcription in the tumor's parent cell line. The expression was evaluated for glucose, insulin, or glucagon in a substrate background of 1 mM lactate. Parallel experiments were carried out in AS-30D hepatoma cells. In contrast to the highly modulated promoter activities observed in AS-30D cells for glucose, insulin, and glucagon, the promoter showed no significant modulations in activity when placed within hepatocytes (Figure 10) and tested with the same modulators. In hepatocytes and in AS-30D cells, the basal activity of the promoter (in 1 mM lactate), as measured by relative light units for the reporter gene, were comparable in magnitude. These results implicate the presence in AS-30D hepatoma cells of one or more transcription factors essential for the expression of the Type II hexokinase gene that are absent in the parental cell line of origin.

EXAMPLE 6: RESPONSE ELEMENTS

Analysis of the 4.3 kbp proximal promoter region of the AS-30D tumor Type II hexokinase revealed both a putative TATA box (AATAA, -30) (Breathnach, et al., (1981), *Annu. Rev. Biochem.*, 50:349-383) and a CAAT box (-85), indicating the precise positioning of transcription initiation for the tumor Type II hexokinase mRNA transcript. This is in contrast to the staggered transcription initiation, and the lack of either a TATA box or a CAAT box, observed for liver glucokinase (Magnuson, et al., (1989), Proc. Natl. Acad. Sci. USA, 86:4838-4832), the principal expressed hexokinase isoform in normal liver cells. Interestingly, response elements for glucose, insulin, glucagon, cAMP, and the phorbol ester TPA were identified for the Type II tumor hexokinase promoter.

Putative consensus sites for Ap-2 (GGCAGCCC, -41), a factor inducible by both protein kinase A and protein kinase C pathways (Faisst, supra; Locker, supra), and for ATF-1 (CCACGTC, -70), which is specifically induced by the protein kinase A pathway, were located immediately upstream of the transcription start site. Since both these sites are located in close proximity to the TATA element and the CAAT

element, further studies will indicate their importance in transcription enhancement. In addition, Ap-2 sites were the most common and ubiquitous elements within the 4.3 kbp promoter (-3850, -2040, -1965, -1500, -1260, -1110, -665, -315). Six putative Ap-1 consensus sites (-3469, -2735, -2320, -1955, -1590, -860) for the complex fos-jun, which is a primary nuclear transducer of the protein kinase C cascade, could be found throughout the 4.3 kbp promoter.

Of the known liver-enriched transcription factors HNF-1, HNF-3, HNF-4, and c/ebp (Lai and Darnell, (1991), *TIBS*, 16:427-430; Lemaigre and Rousseau, (1994), *Biochem. J.*, pgs. 1-14), putative consensus sites could be found for only c/ebp (-4150, -3725, -2550, -1440, -1060, -660, -620, -260). However, several putative sites for the factor HNF-5 (Grange, et al., (1991), *Nucleic Acids Res.*, 19:131-139), which usually binds at sites in close proximity to the response elements for the above-mentioned liver-specific factors, could be found distributed within the promoter (-4160, -3915, -3330, -2200). Therefore, the tumor Type II hexokinase promoter may contain additional consensus sites for the hepatic nuclear factors, or for their oncogenic variants, such as vHNF-1, which replaces HNF-1 in de-differentiated cells (Faisst, supra; Locker, supra). HIF-1 protein (hypoxia-inducible factor) was recently discovered (Semenza, et al., (1995), J. Biol. Chem., 269:23757-23763; Wang and Semenza, (1995), J. Biol. Chem., 270:1230-1237). A response element for HIF-1 was found. As a result, low oxygen concentrations are likely to result also in an activation response via the HIF-1 protein (hypoxia-inducible factor). HIF-1 and glucose response elements are essentially identical. As a result, signal transduction pathways involving glucose, tyrosine kinase, protein kinase A, and protein kinase C are implicated in the transcriptional regulation of tumor Type II hexokinase. Consensus sites for such factors remain to be elucidated by DNA footprinting analysis of the Tumor Type II hexokinase promoter.

Regarding known ubiquitous factors that regulate expression of genes coding for glycolytic and gluconeogenic enzymes (Lemaigre, supra), namely sites for 'CCAAT-box' binding factors, Ocatmer factor, Sp-1, CREB/ATF, Ap-1, b-HLH, and nuclear hormone receptors all of which enhance transcription, putative sites could be

found for Sp-1 (4 sites), CREB/ATF (2 sites), Ap-1 (5 sites), and for steroids (SRE) (5 sites). Also found within the promoter was one putative site (-2955) for factor PPAR, a member of the steroid hormone receptor superfamily, that is thought to play a role in tumor development in liver and in triglyceride and cholesterol homeostasis (Issemann and Green, (1990), Nature, 347:645-650).

Three response elements for the factor p53, a well-known tumor suppessor gene product, were identified in the proximal and distal region of the promoter (-4240, -4195, -1610).

Sites for Pea-3, a factor inducible by TPA, EGF, and the oncoproteins v-src, v-mox, v-raf, and c-Ha-ras (Gutman and Wasylyk, (1990), EMBO, 9:2241-2246) were identified within the distal (4 sites, -3965, -3645, -3625, -3255) and proximal (2 sites, -1415, -1370) regions of the promoter. Within the 4.3 kbp promoter, four Sp-1 binding sites were identified (-3290, -2220, -1110, -55). Since DNA bound Sp-1 factors self-associate, these sites, placed approximately at 1000 bp intervals within the promoter, may bring together the distal promoter segments for enhancement of transcription. Between base pairs of -3811 and -3841, a 31 bp 'GT' repeat was located. This motif, located in the distal region of the tumor Type II hexokinase promoter, is also found within the proximal promoter region of rat pancreatic beta-cell glucokinase (a 33 bp tract) (Magnuson and Shelton, (1989), J. Biol. Chem., 264:15936-15942), as well as within a human glucokinase gene associated satellite repeat DNA sequence (a 31 bp tract) (Tanizawa, et al., (1992), Mol. Endocrinol., 6:1070-1081). Such repetitive purine-pyrimidine DNA segments have potential to form Z-DNA structures, and induce changes in the helicity of adjoining B-DNA.

EXAMPLE 7: INTRODUCTION OF AN ANTISENSE MOLECULE TO REDUCE HEXOKINASE ACTIVITY:

It may be desirable to inhibit the expression of the Type II hexokinase promoter and/or gene because of its association with cancer. Various approaches may be taken to reduce the hexokinase activity in tumor cells. One approach involves the introduction of an RNA molecule (antisense polynucleotide). Antisense technology

involves the juxtaposition of the targeted gene in a reverse orientation behind a suitable promoter, such that an antisense RNA molecule is produced. This antisense construct is then transfected into the engineered cell and, upon its expression, the engineered cell produces a RNA molecule that will bind to, and prevent the processing/translation of RNA produced by the targeted gene, in this case the hexokinase Type II gene.

In order to generate antisense molecules with exact sequence identity to the homologue of hexokinase II being expressed by the tumor cells, the hexokinase variant present in the AS-30D hepatoma cell line will be converted to cDNA by reverse transcribing the mRNA and amplification of the DNA product (Hushes, et al., (1991), J. Biol. Chem., 266:4521-4530). The oligonucleotides used for amplification will be based upon the published sequence of the rat skeletal muscle hexokinase II (Thelen, supra). The oligonucleotides include restriction enzyme recognition sequences at their 5' ends to facilitate directional cloning of the amplified cDNA into the selected vector in an antisense orientation. Because the vector contains both the transcription termination and polyadenylation signal sequences downstream of the cloning cassette, processing of the antisense transcripts should proceed normally. An antisense molecule can be made synthetically in reverse orientation to the sequence provided in Figure 5 or variations of such specific response elements of the sequence. Alternatively, expression constructs may be used which comprise a promoter operably linked to at least 20 nucleotides of the antisense strand of Type II hexokinase cDNA. The expression construct directs the synthesis in a cell of a RNA molecule which is complementary to Type II hexokinase mRNA.

EXAMPLE 8: DRUG SCREENING

The described sequences or variations thereof can be used to screen substances for potential therapeutic agents for disease states, such as cancer and NIDD. A substance which decreases the activity of the Type II hexokinase promoter or production of the gene, is a potential therapeutic agent for cancer treatment. One which increases activity is a potential drug for treating diabetes. Means used to determine amounts of activity are well known in the art, including, but not limited to,

radioactive components, antibodies, etc. For example, use of Type II hexokinase promoter elements driving expression of any reporter gene permits identification of pharmacologic agents capable of depressing or increasing the function of the promoter. Detection of loss of expression of Type II hexokinase or any marker gene by *in situ* hybridization (Baldino, et al., (1989), Methods in Enzymol., 168:761-777; Emson, et al., (1989), Methods in Enzymol., 168:753-61; Harper, et al., (1987), Methods in Enzymol., 151:539-551; Angerer, et al., (1987), Methods in Enzymol., 152:649-661; Wilcox, et al., (1986), Methods in Enzymol., 124:510-533) experiments can identify agents which have the potential to supress the function of the Type II hexokinase promoter. One preferred method for detecting mRNA associated with expression of the cross-reactive protein is *in situ* hybridization to tissue sections, preferably from tumors.

EXAMPLE 9: THERAPEUTIC

The present invention includes methods to treat humans and animals determined to have cells which have an increased rate of glucose utilization over normal cells and/or the capacity to sustain high rates of glycolysis under aerobic (solution or physiological fluid saturated with dissolved oxygen at room temperature (25 degrees Celsius)), hypoxic or under conditions known as hypoxia (low oxygen levels and/or a solution or physiological condition having less saturated conditions as compared to aerobic conditions but not reachning anerobic conditions), or anerobic conditions (solution or physiological fluid having extremely low [near zero] oxygen levels but not hypoxic levels), such as cancer cells. Gene therapy for cancer involves use of hexokinase II transcriptional regulatory regions to drive expression of a toxic gene in tumor cells to kill the tumor cells or inhibit their growth.

Another method of treatment included in the invention is to treat animals determined to have a decreased level or rate of glucose utilization over normal cells and/or an incapacity to sustain high rates of glycolysis under aerobic conditions, such as NIDD. In addition, even normal rates of glucose utilization can be elevated according to the invention in order to lower blood glucose levels. Individuals who

have NIDD may have a defect (mutation) in the insulin receptor, the glucose transporter, or the Type II hexokinase. Gene therapy for NIDD involves introducing to the cell the Type II hexokinase gene with its promoter in an unmethylated form so that it will be expressed.

EXAMPLE 10: HYPOXIA INDUCIBILITY

Within the distal region of the regulatory region of tumor hexokinase II, a potential hypoxia inducible factor (HIF-1) binding motif (CACGTGCT) is present at nucleotides -3765 to -3758. To determine whether the promoter was responsive to hypoxic conditions, we used a promoter-luciferase reporter gene construct, in transient transfection experiments where the tumor cells were maintained in an environment of 1% oxygen. Reporter gene analysis done 24 hrs post-transfection, indicated activation of the type II hexokinase promoter by hypoxia. A 4- to 7-fold activation of the promoter was observed under different substrate backgrounds of pyruvate (1 mM to 10 mM) and glucose (5 mM to 25 mM). This study suggests the involvement of oxygen partial pressure as another step in regulating the transcriptional control of the Type II hexokinase gene, in controlling the glycolytic flux of cancer cells at the first and committed step in glucose catabolism. This is consistent with the observations that most malignant tumor cells survive and proliferate in an oxygen depleted, hypoxic environment. In addition, it is consistentrapidly growing, highly malignant tumor cells, a high rate of glycolysis is maintained irrespective of the in vivo oxygen stress.

EXAMPLE 11: RELATIONSHIP BETWEEN THE TYPE II HEXOKINASE GENE AND P53 PROTEIN OF TUMOR CELLS

In order to elucidate the basis for high expression of the p53 protein in the experimental rat hepatoma cell line AS-30D, preliminary studies were completed in cloning, and sequencing the p53 cDNA. The p53 protein under study was found to be mutated at two positions: 103 (Gly→Ser) and 256 (Glu→Gly). "In vivo" overexpression of the p53 protein was then carried out in AS-30D tumor cells co-transfected with the Type II hexokinase promoter-luciferase reporter gene construct.

The results show that the promoter for the Type II hexokinase gene is positively regulated by the p53 protein. This indicates also that the p53 elements identified on the 4.3 kbp proximal promoter of the Type II hexokinase gene by computer analysis are in fact functionally active.

The cloned p53 cDNA was co-expressed with the Type II hexokinase promoter-luciferase reporter gene construct. A cytomegaloviral (CMV) promoter-vector was used to drive expression of cloned p53 in the tumor cells. Transfected cells were maintained in RPMI-1640 medium (serum-less) containing 1 mM pyruvate. The cells were lysed 20 hrs post-transfection, and luciferase activity measured. The CMV vector lacking a cDNA insert was used in parallel co-transfection experiments as a negative control. As a separate control for the p53 response elements on the hexokinase promoter, a 108 bp minimal promoter construct of the hexokinase gene (denoted Mut C) which contained the TATA element, an ATF and an AP2 element, and the CAAT element, was used. When experiments were repeated in the presence of serum, the induction levels observed for the full length hexokinase promoter were of similar magnitude to that observed under serum-starved conditions, demonstrating the high expression of p53 via the CMV promoter in co-expressed cells. The activation seen for the Mut C construct in the presence of serum is possibly due to activation of the AP2 and ATF elements by the PKA and PKC pathways upon serum induction.

EXAMPLE 12: AMPLIFICATION OF THE GENE ENCODING TYPE II HEXOKINASE IN CANCER CELLS

We have demonstrated by Southern blot analysis and fluorescence in situ hybridization (FISH) that in the rapidly growing rat AS-30D hepatoma cell line, enhanced hexokinase activity is associated with at least a 5-fold amplification of the type II gene relative to normal hepatocytes. The amplified genes are located chromosomally, comprise the whole gene, and most likely are at the site of the resident gene. No rarrangement of the gene could be detected. Therefore, overexpression of hexokinase type II in these cells is based, at least in part, on a stable gene amplification.

Cells and Cell Culture. Clone 9 (CRL 1439), a rat hepatocyte cell line, was obtained from the American Type Culture Collection and grown in RPMI 1640 mediun. AS-30D hepatoma cells were grown in the peritoneal cavity of female Sprague-Dawley rats, harvested and purified as described previously. Hepatocytes were isolated from female Sprague-Dawley rats by collagenase perfusion.

Hexokinase Assay. Hexokinase activity was determined spectrophotometrically on whole cell lysates using a glucose 6-phosphate dehydrogenase coupled assay. Activity is expressed as milliunits (mU) defined as the formation of one nmol NADPH per min.

Southern-blot analysis. High molecular weight DNA was isolated from AS-30D hepatoma cells and hepatocytes as described. DNA (30 μg) was digested with the indicated restriction enzymes. To avoid technical problems resulting from incomplete hydrolysis, digestions were repeated several times with an excess of restriction enzymes. The digested DNA was fractionated on a 1% agarose gel and transferred to nylon membranes (Amersham). Probe labelling, hybridization, and detection were performed with the Fluorescein Gene Images System (Amersham) according to the manufacturer's instructions. Either the full-length cDNA or a 260 bp fragment corresponding to the position -197 to +63 of rat skeletal muscle hexokinase Type II were used as probes.

Fluorescence in situ hybridization. The pUC18 plasmid containing the 3.6 kb cDNA clone of the rat hexokinase (HKII) gene was nick-translated with biotin-14 dATP (BRL, Gaithersburg, MD), with 25% incorporation as determined by tritium tracer incorporation. Slides with chromosome spreads were made from AS-30D hepatoma cells and clone 9 (normal control), harvested by standard cytogenetic techniques. Fluorescence in situ hybridization was performed as described with modifications. Probe mix (2X SSCP, 50% formamide, 10% dextran sulfate, 5 ng/(μl biotinylated probe, and 20 (μg/μl salmon sperm DNA) was denatured at 70°C for 5 min, quickly chilled on ice, placed on slides and hybridized at 37°C overnight. Slides were washed in 50% formamide/2XSSC at 43°C for 20 minutes, and 2 changes of 2XSSC at 37°C for 5 min each. Biotinylated probe was detected with FITC-avidin and amplified with

biotinylated anti-avidin, using reagents from "in situ hybridization kit" (Oncor Inc., Gaithersburg, MD), following manufacturer's instructions.

Preliminary Southern-blot analysis using digested genomic DNA from hepatocytes and AS-30D hepatoma cells revealed that the hexokinase Type II probe hybridized with much greater intensity to the hepatoma DNA than to the hepatocyte DNA. To estimate the differences in hybridization intensities we performed a dilution experiment. The hybridization signals with different amounts of EcoR I/Xba I digested AS-30D hepatoma genomic DNA were compared to the signal obtained with 30 μg of DNA isolated from hepatocytes (Fig. 1). The blots were probed with two different probes specific for the hexokinase Type II gene (Fig. 1A and 1B). The intensities of the resulting bands indicate that 3-6 μg hepatoma DNA were equivalent to 30 μg hepatocyte DNA. From this experiment we estimated that AS-30D hepatoma cells contain approximately 5-10 fold more copies of the hexokinase Type II gene than normal hepatocytes. In addition it is clear from Fig. 1A that the signal intensities of all Type II hexokinase related bands obtained with AS-30D hepatoma DNA are the same. This indicates that the amplification extends to the whole coding region of the hexokinase gene. Moreover, when the membranes were probed again with DNA fragments specific for the 5'-flanking region of the hexokinase gene similar results were obtained (data not shown). Thus, the amplified unit in AS-30D hepatoma cells also includes the promotor region of the hexokinase Type II gene. Densitometric quantification of autoradiograms made from different Southern-blots confirmed the data obtained in the dilution experiment, and a factor of approximately 5 was calculated for the amplification. Additional support for the hexokinase Type II gene amplification in AS-30D hepatoma cells came from experiments searching for the hexokinase Type II promotor region in these cells and in hepatocytes. Thus, 6 positive plaques were obtained when 5 x 105 plaques were screened from an AS-30D hepatoma genomic library, whereas only 2 positives were found in 2.5 x 106 plaques of a normal liver library. Taking into consideration that the liver library had been prior amplified the estimated factor for amplification is near 6, in accordance with results from Southern-blot analysis.

Instability of the genome is a well-known phenomenon of transformed cells and amplification is a frequently observed mechanism for the overexpression of oncogenes including N-myc and the epidermal growth factor receptor gene. It is well known that a strong relationship frequently exists between a gene that is amplified and cell growth. The amplification of the hexokinase Type II gene is consistent with this relationship as the role of this critical metabolic enzyme is to provide cells with both energy and precursors for nucleotide and lipid biosynthesis. In a recent report, we provided evidence that increased expression of one or more transcription factors is involved in the elevated production of hexokinase Type II in AS-30D hepatoma cells. Work presented here suggests that amplification of the gene for the same enzyme may play a role as well.

Southern-blot analysis displayed some faint restriction fragments with the hepatocyte DNA which were not observed in the AS-30D hepatoma DNA. As the restriction enzymes used, EcoR I and Xba I, are both sensitive to methylation of their recognition sequence, this raises the possibility that methylation differences exist within the hexokinase Type II gene in normal hepatocytes and AS-30D hepatoma cells. Several studies reviewed in have demonstrated that DNA methylation plays a role in gene regulation. Therefore, methylation could be involved in differential expression of hexokinase Type II in normal and tumor cells. Further experiments to test this hypothesis are in progress.

For some oncogenes it is well known that amplification is accompanied by recombination and rearrangement of the gene locus. To look for structural differences in the hexokinase Type II gene locus in normal and AS-30D hepatoma cells restriction fragment length polymorphism (RFLP) analysis was carried out. To circumvent problems due to methylation differences of normal and tumor DNA, methylation insensitive restriction enzymes (Rsa I, Nde I, Hinc III) were used. For each enzyme the same restriction fragment pattern is observed in both hepatocyte and AS-30D hepatoma DNA. Thus, no macroscopic rearrangement of the hexokinase gene is seen at this level of resolution. Also, this result renders it unlikely that a translocation of the hexokinase gene locus has occured in AS-30D hepatoma cells. Therefore, the

amplification described above appears to occur at the site of the resident gene, and the possibility that the hexokinase Type II gene in AS-30D hepatoma cells has come under the control of different regulatory sequences through translocation seems remote.

To obtain additional support for the amplification and localization of the hexokinase Type II gene, in situ hybridization experiments were performed. Because primary hepatocytes divide very rarely and rapidly dedifferentiate, we used clone 9 (CRL 1439), a non-tumorigenic, normal liver cell line as a control for in situ hybridization. These cells exhibit no detectable hexokinase activity in contrast to AS-30D hepatoma cells where the activity is 762 mU/mg. The liver homogenate, which in addition to hepatocytes contains other cell types, exhibits a low but detectable hexokinase activity. In situ hybridizations using the hexokinase Type II cDNA as probe revealed that in AS-30D hepatoma cells a signal could be readily detected in every metaphase (20/20) and interphase cell. Occasional (4/20) tetraploid cells which were observed in the AS-30D hepatoma cell population showed a hybridization signal on two chromosomes indicating that the gene was amplified before the chromosomes were duplicated. The single positive chromosome seen in the AS-30D sample most likely represents the amplification site on one chromosome homolog only, but the loss of the other homologous chromosome cannot be ruled out. In contrast, in clone 9, no interphase signals were seen and only 1 of 20 metaphase cells showed a faint specific signal.

Because the probe used for in situ experiments was rather small (3.6 kbp), genes with a low copy number cannot be easily detected with this size probe. This confirms again that the copy number of the hexokinase Type II gene is much more abundant in AS-30D hepatoma cells than in control cells. Although FISH does not allow exact quantitation of the amplification, it is consistent with at least a 5-fold increase in copy number. Moreover, the amplified sequence was localized to a single chromosome in AS-30D hepatoma cells suggesting that the amplification is present on only one of the two homologous chromosomes, a finding not uncommon for amplified genes (20). Chromosomally localized gene amplification represents one of the more stable forms of amplified genes. Stable retention of amplified genes and their passage

to daughter progeny are ensured only when such genes are integrated within a chromosome. Unstable amplified genes which are very common in transformed cells are characteristically associated with extrachromosomal elements called double minutes. However, double minutes were never observed in our studies of AS-30 hepatoma cells.

In summary, results reported here provide for the first time evidence that a hexokinase gene (Type II) is amplified in a tumor cell line exhibiting a high glucose catabolic phenotype. This amplification is stable, not associated with a rearrangement of the hexokinase gene locus, and probably occurs at the site of the resident gene.

The invention has been described with reference to the presently preferred embodiments, it should be understood that various modifications can be made without departing from the spirit of the invention. All references and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: PEDERSEN, PETER L.
MATHUPALA, SAROJ P.
REMPEL, ANNETTE

(ii) TITLE OF INVENTION: TUMOR TYPEII HEXOKINASE TRANSCRIPTION REGULATORY REGIONS

(iii) NUMBER OF SEQUENCES: 3

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: BANNER & ALLEGRETTI, LTD.
(B) STREET: 1001 G STREET, ELEVENTH FLOOR
(C) CITY: WASHINGTON DC
(E) COUNTRY: USA
(F) ZIP: 20001

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 60/001,199
(B) FILING DATE: 14-JUL-1995
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: KAGAN, SARAH A.
(B) REGISTRATION NUMBER: 32,141
(C) REFERENCE/DOCKET NUMBER: 1107.57886

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 202-508-9100
(B) TELEFAX: 202-508-9299

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5150 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Rattus rattus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TCTAGAGCTC GCGGCCGCGA GCTCTAATAC GACTCACTAT AGGGCGTCGA CTCGATCCAA 60

CTGGCCTAGA ACTCACAGCC ATCCTCTTGC CTCTACCTAT GGAGTGTGGA GATTAAAGGC 120

ATGTTCTACC ATGTCTTAAT TTTAAAATAC CTATGGAGTG TGGAGATTAA AGGCATGTTC 180

TACCATGTCT TAATTTTAAA ATAGGAATAT TTGTGGATTG AGGTCTTGAG CAAAATAAGA 240

TTTTTCCCAA GAGAGTTTCC TGAAGCCTAA GTAGACTCAG GTCCTTCTCA TGCAGGGCCA 300

ATCTAGGGCC AGGAGCAGGA CCAACTGGTG TGAAATCAGA AAGATGGTAC TCATAGCTAT 360

TAGTCCATCT CTGGTTGACA CTCCCAGACT CCCCTACATC TCAAGACACA GACATACGTG 420

GCTTTTTATG AATCCATTTT TCTGGTCTGT ATTATTTGTT CTGTGTGTTA ATTTTATGTC 480

TTAAACTAAA CAGAAATCCT TTAAGGAAAG AACACCCCGC CCCTCTCCTG TGTGTGTGTG 540

TGTGTGTGTG TGTGTGTGTA CACGTCTGTG TGAGTATCTC GCACCCTGTA AAGGGCTTAA 600

TAAACACGTG CTGATTGATT CCCCTCCTGG AATGTGAATG TGGACTGCCA ATCTGCCAGT 660

CTACAATGTG TGTGCCTGTA TGTGCTCATG GGGAGAGAGA GGGAGAAATA AAATAGACTC 720

TAAGGAAGAA TCTTGAGGAC AGGAAAGTCA GAGCTACACA CCTCACTTTT GAGTGGGTAG 780

CTGTCCCCTG ATTTGACACA TACAGATGGG TTAGGGGATA TCACTGTACT CACTCCAGCC 840

ACCTCCCAGG GTTACTGGGA ACTCTGTGAG AGATCATCCC ATAAAGTACC CTGTGAACAT 900

GAGTTAGTCC TCATAAAGTG GGACCAGAAA AGAGAATGGA GAATGGAGCT GAAGTGTGTG 960

TGCAAGTAAG TGTGTGTGAG ATCCAGCTAA TTGGACTCAG CTGATGGAGT GCTTGCCTAG 1020

CACGCATGAA TCCTCATGTT TGCCTCTGAT CGCAAGACCT GAAAAAAAAA AAAAATAGGC 1080

GAGGTAGACA GTGCCTGTAA CCTCAGCGCT GAGGAAGTGG AGGCCGGAGG ATGGGAAGCT 1140

CAAGACTGTC CTTGGTTGCA TGTTTAGTTA GAGGCCATCT TGGGCCACAT GATCCTGTCC 1200

CAAAATAAAC AAAGGAATAC AATTAGTCCA TAGGGAGGAG ATCATAGTTG ACCTGACCCC 1260

ACTGATTTTG ATCTTAGTTG TCTGAGGGAA ATTATTTTAT ATACTGATTT AACTATCGGT 1320

TTTTTAAGTG TCTCAAAATG TTTTTATTTC ATGTACACCC TTATTGGGTG TGCATATGCA 1380

TGTAGGTACA CACATGCCAT GGTAAGTGTG TAAAGGTCAG AAGTCAATTT TCTTGAGTTG 1440

ATTCTCTCCT GTGACCACAT GGGTCCTAGG GTTCACCTCA AGTAGTCGGG GTTCAGCGAC 1500

AAGCGCCTTT ACCCACCGAG CCACCTTGCC AGCACCCGAA GTGTTTCCAG AAAGGTCTTT 1560

TTTTTTTTCT CTTTGCTGCT TACTTTTAAC CTATGCCATC AATTCTGCCT CAGACTTCTG 1620

AACACCTAAA GCCTTAATCA GCCTCTGTGC CTCACCCTTG TCTCACTCCA GCCTTTATCT 1680

TATCTGGGAG TTCCTGTCTC TTCTCCTTCA GGCCGGGTCC TTTCCTCCCA TTCATGTGGA 1740

GAGCAGCTTT TGTCCTACAA AAGCTTTAAG CATCTCAGAG TCTGTGTCAG AAAGAGAGGA 1800

GCTGGCTTAT GAGGCTGTTG CAATTGGGTG AAAGACACTG GTGAACTGTG AGGCAGACCA 1860

ATGGGAAGGG TTTGAGAACT AATATAGAAA ATGAAAGTCT CTCCTTTGTG TCGTATAATC 1920

ATATGTGACA TCACTAAATC ATCTACTAAC TTACACAATA AATACCTACA TGGTGCCTAC 1980

CATGTGATAG AGCGCCCCCA CACGAGGTAC TGCAGATAGA AGGGAATGAT ATAGACGCAG 2040

ATGCTTATTC AGACAGGTAG GACAGAATGG ATATAACACT TAGAAAAGGA CCCGGGTGTG 2100

GTGGCACGGT GGCACATACC TTAGATCCCA CCACCCGGGT GTTGGGGCTG AGGCAGATGA 2160

CTTTTGTTTG TTTTTTTGTT TTCAGTTTTG CTTTTTTTCA AGACAGGGTC TCTCCGTGCA 2220

GCACTGCCTG TCCTGGAACT CGCTTTGCTG GCCACGTTTG TGGCCTTGAA CTCACAAAGT 2280

GCTGGTGCCT GGCTGTAAAA TTAATTTCTC TCTCCCTCTC TCCCCCCCTC CCCCACCTCT 2340

CTCGCTACTT GCTTGGTAGA CCAGACTGGC TTCGAACTCA GAGATTTGCC TGCGTTTGCC 2400

GCCCAAGGGC TGTGATTAAA GGTATGTGCC ACCATGTCCA GCCTTAAAAA TTACTTCTAA 2460

TAGTCATTCT TAGGAGTTTG GATTTTATTT GAAGATAAGA AAACAATAAT GGTTTTAAGA 2520

CTCTTCCCCC CCAAAAAGAC AGTTTGGTAT ATATCTATCA ATCAATCTAA TCTTATCTCC 2580

TGCCTGCCTG CGTATCTATC TATCTATCTA TCTATCTATC TATCTATCTA TCTATCCATC 2640

CATCCAAGGT CTCATGCTTA CCAAGTTGGG CTTGAACTCC TGACTCTTCT GTCTCCACCT 2700

CTGGAGCACT GGGACTACAA ATTTGTGCCA CCCCACAAAG CACTGGCTTG TATTTTAAAC 2760

AAGTCTCTTT AGCTCTTGAG TAAGAGGGTT CATGGTGGTC AAACTAGAGG TAGCTAAAAA 2820

TGGCAGCTAA GTGACATTAC ACGGACTCGG GTGGAGTCAT CGATGGCCTG GCCATGAGGG 2880

TCTGGCCTTT TTGATTTGCA GTTAGACTAA CTGCTCCCCG ATGGAGTGGA TAGTTGTAAG 2940

AGCAGGTGGC AGGAAGACAC CATGGATGGT GATGTCATTT GTGGAGACAA CTGGTAAAGG 3000

AAAAAAAAAA CCTAAGAAGT TCAGCTTAGT ACATGTTAAG TGTGAGGTGC TAGTCACCTC 3060

TGCTGGGATG ACAAACCATA GCTGGCTAAG AAAGTTAGAA GCCCTGGGGG AAAGTTCTTG 3120

CTTTGTAGAT TAGGTTTGTG AGTACCTGTG TGCAGATGGT GTAACTGCTG TCGACAGTGC 3180

TGGGAGGAAT CGCCCATCGA GGGAATAGAT GAACCACACC AGAGAACATG GTAGAAGCGG 3240

CCCAGCAGAG CAACGTGGGC TGGGGTGTAC TTCAGTCGGC AGAGTGCTTT GATCTCCAGT 3300

AGTGGCCCAT GCCACTCCAG AGGTGGGGAG AGAGCTTGGG GAGCGAGACT GTTTGGAAAT 3360

GGTAGGCCCT GTCTTCTTTC CATGTAACTT TCCAACTCCC AGGTTTCCCA TTCTCCACCA 3420

GCAACAACCT CATGCCATTT GAGGTGCTAC TTCAATATCG CTGGCGTCTA CTCATCTATG 3480

TGAACTTAAG AGTCTACTCA TCTATGTGAC TAAGAGTCTG GTGTCAGGCG TGAGCTGAGG 3540

TAGAGGTGGG CTCTTCTCAG CCTCTATAAA CCAATTCACA CCACTTGAGC CAAGCAGTTA 3600

CACATGCACT TTCTCCTCCC GCCTATCAGT CCTAGCTCCT GACAAGGTTT CTCTCCAGCC 3660

TTTTACTTTC CTGGCTTCAA GAAAGGCGGG ATAATATACC AGGGTGGGGA GATTGCATTT 3720

CAGAGTGAGA CGTGTTCTGT CTTCACCTAC CACTTGTTGG CGATGTGACC TTGGGCAAAG 3780

CTCATTAACA GCACAGTGCC TAGTTCCCTA ATTTGTAAAA CATATGCTAT AGGTGTGACG 3840

ATTACGAAGG GCTGACTTTT GTAATGGCTT TGCTTCAGGG ATCTGCAGAC TCGTTGAGCC 3900

ACAATTAGGA TGAGAATCAA GGTGCTTCAG ACTTGTGACA GGGCACTGGC GGCCCCTCAC 3960

ATGATCCTCA GATACCAGAT TGTGGCGTGT GCTGCTAGGA TCACTTGTCT TTCCAGTCTC 4020

CCAACATCTC TTGGGTCCGT GATCACGCGC CCCCCACCCG AAGCCCAGCC TGACGCGGCG 4080

GTGGCTCATG CGCCCTGGAG TCCCGGGCTC TAGCCACGGA ACACACGTCC CAACTCTGGC 4140

GCCCGGCTCC GCCCCTAGCC TCGGGCGCGT CTCTCCCGCC GCCTGCTTGG GTGCTGGAGC 4200

AGCCGCGCCC GCGGGCTCTG GGCGCTGATT GGCTGTGGAC TGCGGGCGGG CAGCCGGAGA 4260

GCGTACACAC CCTCTTCCCG CAGCCAATGA GCGCGCCCAC GTCACTGTCT TGGGCGGCCC 4320

AAAGAGCCGG CAGCCCCTCA ATAAGCCACA TTGTTGCACC AACTCCAGTG CTAGAGTCTC 4380

AGGACACCAC AGGCTACACG GAGTTATCCC GCTTAGGAGA CCCGAAGGCA GGAGCATCAC 4440

TCCAGTGACT CTGATAAGGT GCGATCGCCC GAGAGGAACA GAACTGTCAT TTTTGCGAAG 4500

TTGAGCCTTA CGGATCCCGT GGGCGAAGTT AGCGACGGGA CGCTGAGCAA CTAGACCGGA 4560

CGGCAGGAGT GAGACTTAGG TGCCTTCTAG TAGTTGTGAC TTAAAAAAAA AAAAAAAAGG 4620

AAAAGAAAAA AGGAGGAAAA CCTGTTTCTG GAAACGCGAG GCCCTCAGCT GGTGAGCCAT 4680

CGTGGTTAAG CTTCTTTGTG TGGCTCCTGG AGTCTCCGAT CCCAGCCGGA CACCCGGGCC 4740

TGGTTTCAAA GCGGTCGAAC TGCTCTGCCC GCTCCACCGG TAGCGCTCGA GCCTCGGTTT 4800

CTCTACTCGA CCCCGACTCG CCGCAGCAGG ATGATCGCCT CGCATATGAT CGCCTGCTTA 4860

TTCACGGAGC TCAACCAAAA CCAAGTGCAG AAGGTAAGTC GGCACGGGCG GGAGCTGCTG 4920

GCTCGCTTCG GACCAAGTTG CGTGCTCTCC GGGAATCTGG AGCACGCAGA GGACCTGCTT 4980

CCTCCTCCGG GGCTGGGGAC GTGGAACCAG TCTGAGTAGC TGGGAAAGTC CTGAGCGCCA 5040

GAAACCACGT CTGCTAGGCA CCCTCGTGGC CCGGCCGCGC ATCACCGATA CTCCCACTTT 5100

CCCGGGATCC GCGAGCATCC TCCCCACCCT TAAAGCCCCT AATTTCTAGA 5150

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2771 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Rattus rattus

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 18..2771

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
CGACTCGCCG CAGCAGG ATG ATC GCC TCG CAT ATG ATC GCC TGC TTA TTC          50
                   Met Ile Ala Ser His Met Ile Ala Cys Leu Phe
                     1               5                  10

ACG GAG CTC AAC CAA AAC CAA GTG CAG AAG GTT GAC CAA TTT CTC TAC          98
Thr Glu Leu Asn Gln Asn Gln Val Gln Lys Val Asp Gln Phe Leu Tyr
              15                  20                  25

CAC ATG CGT CTC TCA GAT GAG ACC CTT CTG GAG ATT TCT AGG CGG TTC         146
His Met Arg Leu Ser Asp Glu Thr Leu Leu Glu Ile Ser Arg Arg Phe
          30                  35                  40

CGG AAG GAG ATG GAG AAA GGG CTA GGA GCT ACC ACG CAC CCT ACA GCA         194
Arg Lys Glu Met Glu Lys Gly Leu Gly Ala Thr Thr His Pro Thr Ala
      45                  50                  55

GCT GTG AAA ATG TTG CCT ACC TTT GTG AGG TCA ACT CCG GAT GGG ACA         242
Ala Val Lys Met Leu Pro Thr Phe Val Arg Ser Thr Pro Asp Gly Thr
  60                  65                  70                  75

GAA CAT GGG GAG TTC CTG GCT CTG GAT CTT GGA GGA ACC AAC TTC CGT         290
Glu His Gly Glu Phe Leu Ala Leu Asp Leu Gly Gly Thr Asn Phe Arg
                  80                  85                  90

GTG CTC CGA GTA AGG GTG ACG GAC AAT GGC CTC CAG AGA GTG GAG ATG         338
Val Leu Arg Val Arg Val Thr Asp Asn Gly Leu Gln Arg Val Glu Met
              95                  100                 105

GAG AAC CAG ATC TAC GCC ATC CTT GAG GAC ATC ATG CGG GGC AGT GGA         386
Glu Asn Gln Ile Tyr Ala Ile Leu Glu Asp Ile Met Arg Gly Ser Gly
          110                 115                 120

ACC CAG CTG TTT GAC CAC ATC GCC GAA TGC CTG GCC AAC TTC ATG GAC         434
Thr Gln Leu Phe Asp His Ile Ala Glu Cys Leu Ala Asn Phe Met Asp
      125                 130                 135

AAG CTA CAA ATC AAA GAG AAG AAG CTC CCT CTG GGT TTC ACC TTC TCG         482
Lys Leu Gln Ile Lys Glu Lys Lys Leu Pro Leu Gly Phe Thr Phe Ser
140                 145                 150                 155
```

```
TTC CCC TGC CAC CAG ACA AAA CTG GAT GAG AGT TTT TTG GTC TCG TGG        530
Phe Pro Cys His Gln Thr Lys Leu Asp Glu Ser Phe Leu Val Ser Trp
                160                 165                 170

ACT AAG GGG TTC AAG TCC AGT GGC GTG GAA GGC AGA GAT GTG GTG GAC        578
Thr Lys Gly Phe Lys Ser Ser Gly Val Glu Gly Arg Asp Val Val Asp
            175                 180                 185

CTG ATC CGG AAG GTT ATC CAG CGC AGA GGG GAC TTT GAC ATT GAC ATT        626
Leu Ile Arg Lys Val Ile Gln Arg Arg Gly Asp Phe Asp Ile Asp Ile
        190                 195                 200

GTG GCC GTG GTG AAT GAC ACA GTT GGG ACC ATG ATG ACT TGT GGC TAT        674
Val Ala Val Val Asn Asp Thr Val Gly Thr Met Met Thr Cys Gly Tyr
    205                 210                 215

GAT GAT CAG AAC TGC GAG ATT GGT CTC ATT GTG GGC ACT GGC AGC AAC        722
Asp Asp Gln Asn Cys Glu Ile Gly Leu Ile Val Gly Thr Gly Ser Asn
220                 225                 230                 235

GCC TGC TAC ATG GAG GAA ATG CGT CAT ATT GAC ATG GTG GAG GGA GAT        770
Ala Cys Tyr Met Glu Glu Met Arg His Ile Asp Met Val Glu Gly Asp
                240                 245                 250

GAG GGG CGC ATG TGC ATC AAC ATG GAG TGG GGA GCC TTT GGG GAC GAC        818
Glu Gly Arg Met Cys Ile Asn Met Glu Trp Gly Ala Phe Gly Asp Asp
            255                 260                 265

GGT ACA CTC AAT GAC ATC CGA ACC GAG TTT GAC CGA GAG ATC GAC ATG        866
Gly Thr Leu Asn Asp Ile Arg Thr Glu Phe Asp Arg Glu Ile Asp Met
        270                 275                 280

GGC TCG CTG AAC CCT GGG AAG CAG CTG TTT GAG AAG ATG ATT AGC GGG        914
Gly Ser Leu Asn Pro Gly Lys Gln Leu Phe Glu Lys Met Ile Ser Gly
    285                 290                 295

ATG TAC ATG GGG GAG CTG GTC AGG CTC ATC CTG GTG AAG ATG GCC AAG        962
Met Tyr Met Gly Glu Leu Val Arg Leu Ile Leu Val Lys Met Ala Lys
300                 305                 310                 315

GCA GAG CTG TTG TTC CAA GGG AAA CTC AGC CCA GAA CTC CTT ACC ACT       1010
Ala Glu Leu Leu Phe Gln Gly Lys Leu Ser Pro Glu Leu Leu Thr Thr
```

```
                    320                 325                 330

GGC TCC TTC GAG ACC AAA GAT GTC TCG GAT ATT GAA GAG GAT AAG GAT      1058
Gly Ser Phe Glu Thr Lys Asp Val Ser Asp Ile Glu Glu Asp Lys Asp
            335                 340                 345

GGA ATC GAG AAG GCC TAC CAA ATC CTG ATG CGC CTG GGT CTG AAT CCA      1106
Gly Ile Glu Lys Ala Tyr Gln Ile Leu Met Arg Leu Gly Leu Asn Pro
        350                 355                 360

TTG CAG GAG GAT TGT GTG GCC ACG CAC CGA ATC TGC CAG ATT GTG TCC      1154
Leu Gln Glu Asp Cys Val Ala Thr His Arg Ile Cys Gln Ile Val Ser
    365                 370                 375

ACG CGC TCG GCC AGT CTG TGC GCA GCC ACC CTG GCC GCG GTG CTG TGG      1202
Thr Arg Ser Ala Ser Leu Cys Ala Ala Thr Leu Ala Ala Val Leu Trp
380                 385                 390                 395

CGA ATC AAA GAG AAC AAG GGC GAG GAG CGA CTT CGC TCC ACC ATC GGT      1250
Arg Ile Lys Glu Asn Lys Gly Glu Glu Arg Leu Arg Ser Thr Ile Gly
                400                 405                 410

GTC GAT GGC TCC GTC TAC AAG AAA CAT CCC CAT TTT GCC AAG CGT CTC      1298
Val Asp Gly Ser Val Tyr Lys Lys His Pro His Phe Ala Lys Arg Leu
            415                 420                 425

CAT AAG GCA GTG AGG AGG CTG GTG CCC GAC TGT GAT GTC CGC TTC CTC      1346
His Lys Ala Val Arg Arg Leu Val Pro Asp Cys Asp Val Arg Phe Leu
        430                 435                 440

CGC TCT GAG GAT GGC AGC GGC AAG GGG GCT GCT ATG GTG ACG GCG GTG      1394
Arg Ser Glu Asp Gly Ser Gly Lys Gly Ala Ala Met Val Thr Ala Val
    445                 450                 455

GCT TAC CGT CTG GCT GAC CAA CAC CGG GCC CGC CAG AAG ACC CTG GAG      1442
Ala Tyr Arg Leu Ala Asp Gln His Arg Ala Arg Gln Lys Thr Leu Glu
460                 465                 470                 475

TCT CTG AAG CTG AGC CAC GAG CAG CTT CTG GAG GTT AAG AGA AGA ATG      1490
Ser Leu Lys Leu Ser His Glu Gln Leu Leu Glu Val Lys Arg Arg Met
                480                 485                 490
```

```
AAG GTG GAA ATG GAG CAG GGT CTG AGC AAG GAG ACG CAT GCG GTC GCC      1538
Lys Val Glu Met Glu Gln Gly Leu Ser Lys Glu Thr His Ala Val Ala
            495                 500                 505

CCT GTG AAG ATG CTG CCC ACT TAC GTG TGT GCC ACT CCA GAT GGC ACA      1586
Pro Val Lys Met Leu Pro Thr Tyr Val Cys Ala Thr Pro Asp Gly Thr
        510                 515                 520

GAG AAA GGA GAC TTC TTG GCC TTG GAT CTT GGA GGA ACA AAC TTC CGG      1634
Glu Lys Gly Asp Phe Leu Ala Leu Asp Leu Gly Gly Thr Asn Phe Arg
    525                 530                 535

GTC CTG CTG GTG CGT GTG CGT AAT GGC AAG CGG AGG GGC GTG GAG ATG      1682
Val Leu Leu Val Arg Val Arg Asn Gly Lys Arg Arg Gly Val Glu Met
540                 545                 550                 555

CAT AAC AAG ATC TAC TCC ATC CCA CAG GAG GTT ATG CAT GGC ACT GGG      1730
His Asn Lys Ile Tyr Ser Ile Pro Gln Glu Val Met His Gly Thr Gly
                560                 565                 570

GAA GAG CTC TTC GAC CAC ATT GTC CAG TGC ATT GCG GAC TTC CTG GAG      1778
Glu Glu Leu Phe Asp His Ile Val Gln Cys Ile Ala Asp Phe Leu Glu
            575                 580                 585

TAC ATG GGC ATG AAG GGC GTG TCC CTG CCT TTG GGT TTC ACA TTC TCC      1826
Tyr Met Gly Met Lys Gly Val Ser Leu Pro Leu Gly Phe Thr Phe Ser
        590                 595                 600

TTC CCT TGC CAG CAG AAC AGC CTA GAC CAG AGC ATC CTC CTC AAG TGG      1874
Phe Pro Cys Gln Gln Asn Ser Leu Asp Gln Ser Ile Leu Leu Lys Trp
    605                 610                 615

ACA AAG GGA TTC AAG GCA TCT GGC TGC GAG GGT GAG GAT GTG GTC ACC      1922
Thr Lys Gly Phe Lys Ala Ser Gly Cys Glu Gly Glu Asp Val Val Thr
620                 625                 630                 635

TTG CTG AAG GAA GCG ATT CAC CGG CGA GAG GAG TTT GAC CTG GAT GTG      1970
Leu Leu Lys Glu Ala Ile His Arg Arg Glu Glu Phe Asp Leu Asp Val
                640                 645                 650
```

GTT GCC GTG GTG AAT GAC ACA GTT GGG ACT ATG ATG ACT TGT GGC TAC 2018
Val Ala Val Val Asn Asp Thr Val Gly Thr Met Met Thr Cys Gly Tyr
655 660 665

GAA GAC CCT CAC TGT GAA GTT GGC CTC ATT GTT GGC ACC GGA AGC AAC 2066
Glu Asp Pro His Cys Glu Val Gly Leu Ile Val Gly Thr Gly Ser Asn
670 675 680

GCC TGC TAC ATG GAA GAG ATG CGT AAT GTG GAG CTG GTG GAC GGA GAG 2114
Ala Cys Tyr Met Glu Glu Met Arg Asn Val Glu Leu Val Asp Gly Glu
685 690 695

GAG GGA CGG ATG TGT GTC AAC ATG GAG TGG GGA GCA TTT GGG GAC AAT 2162
Glu Gly Arg Met Cys Val Asn Met Glu Trp Gly Ala Phe Gly Asp Asn
700 705 710 715

GGC TGC CTG GAT GAC TTG CGG ACC GTG TTT GAT GTT GCT GTG GAT GAG 2210
Gly Cys Leu Asp Asp Leu Arg Thr Val Phe Asp Val Ala Val Asp Glu
720 725 730

CTT TCT CTC AAC CCT GGC AAA CAG AGG TTC GAG AAG ATG ATC AGC GGC 2258
Leu Ser Leu Asn Pro Gly Lys Gln Arg Phe Glu Lys Met Ile Ser Gly
735 740 745

ATG TAC TTG GGA GAG ATT GTG CGC AAC ATT CTC ATC GAT TTC ACG AAG 2306
Met Tyr Leu Gly Glu Ile Val Arg Asn Ile Leu Ile Asp Phe Thr Lys
750 755 760

CGG GGG CTG CTC TTC CGA GGC CGC ATC TCA GAG CGC CTC AAG ACA AGG 2354
Arg Gly Leu Leu Phe Arg Gly Arg Ile Ser Glu Arg Leu Lys Thr Arg
765 770 775

GGA ATC TCT GAA ACT AAG TTC CTG TCT CAG ATA GAG AGC GAC TGC CTA 2402
Gly Ile Ser Glu Thr Lys Phe Leu Ser Gln Ile Glu Ser Asp Cys Leu
780 785 790 795

GCC CTG CTA CAG GTT CGT GCC ATC CTG CGC CAC CTA GGG CTG GAG AGC 2450
Ala Leu Leu Gln Val Arg Ala Ile Leu Arg His Leu Gly Leu Glu Ser
800 805 810

```
ACG TGC GAT GAC AGC ATC ATC GTG AAG GAG GTG TGC ACT GTG GTT GCC      2498
Thr Cys Asp Asp Ser Ile Ile Val Lys Glu Val Cys Thr Val Val Ala
            815                 820                 825

CGG CGC GCT GCA CAG CTC TGT GGC GCA GGC ATG GCC GCC GTA GTG GAC      2546
Arg Arg Ala Ala Gln Leu Cys Gly Ala Gly Met Ala Ala Val Val Asp
        830                 835                 840

AAG ATA AGA GAG AAC CGT GGG CTG GAC AAC CCC AAA GTG ACA GTG GGC      2594
Lys Ile Arg Glu Asn Arg Gly Leu Asp Asn Pro Lys Val Thr Val Gly
    845                 850                 855

GTG GAC GGG ACT CTG TAT AAG CTT CAT CCT CAC TTT GCC AAG GTC ATG      2642
Val Asp Gly Thr Leu Tyr Lys Leu His Pro His Phe Ala Lys Val Met
860                 865                 870                 875

CAT GAG ACG GTG AGA GAT CTG GCT CCG AAA TGT GAC GTG TCC TTC CTG      2690
His Glu Thr Val Arg Asp Leu Ala Pro Lys Cys Asp Val Ser Phe Leu
                880                 885                 890

GAA TCC GAG GAC GGC AGT GGG AAG GGA GCA GCT CTC ATC ACT GCC GTG      2738
Glu Ser Glu Asp Gly Ser Gly Lys Gly Ala Ala Leu Ile Thr Ala Val
            895                 900                 905

GCC TGC CGC ATC CGG GAG GCT GGG CAG AGA TA                          2771
Ala Cys Arg Ile Arg Glu Ala Gly Gln Arg
        910                 915
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 917 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Met Ile Ala Ser His Met Ile Ala Cys Leu Phe Thr Glu Leu Asn Gln
 1               5                  10                  15
```

Asn Gln Val Gln Lys Val Asp Gln Phe Leu Tyr His Met Arg Leu Ser
20 25 30

Asp Glu Thr Leu Leu Glu Ile Ser Arg Arg Phe Arg Lys Glu Met Glu
35 40 45

Lys Gly Leu Gly Ala Thr Thr His Pro Thr Ala Ala Val Lys Met Leu
3 50 55 60

Pro Thr Phe Val Arg Ser Thr Pro Asp Gly Thr Glu His Gly Glu Phe
65 70 75 80

Leu Ala Leu Asp Leu Gly Gly Thr Asn Phe Arg Val Leu Arg Val Arg
85 90 95

Val Thr Asp Asn Gly Leu Gln Arg Val Glu Met Glu Asn Gln Ile Tyr
100 105 110

Ala Ile Leu Glu Asp Ile Met Arg Gly Ser Gly Thr Gln Leu Phe Asp
115 120 125

His Ile Ala Glu Cys Leu Ala Asn Phe Met Asp Lys Leu Gln Ile Lys
130 135 140

Glu Lys Lys Leu Pro Leu Gly Phe Thr Phe Ser Phe Pro Cys His Gln
145 150 155 160

Thr Lys Leu Asp Glu Ser Phe Leu Val Ser Trp Thr Lys Gly Phe Lys
165 170 175

Ser Ser Gly Val Glu Gly Arg Asp Val Val Asp Leu Ile Arg Lys Val
180 185 190

Ile Gln Arg Arg Gly Asp Phe Asp Ile Asp Ile Val Ala Val Val Asn
195 200 205

Asp Thr Val Gly Thr Met Met Thr Cys Gly Tyr Asp Asp Gln Asn Cys
210 215 220

Glu Ile Gly Leu Ile Val Gly Thr Gly Ser Asn Ala Cys Tyr Met Glu
225 230 235 240

Glu Met Arg His Ile Asp Met Val Glu Gly Asp Glu Gly Arg Met Cys
245 250 255

Ile Asn Met Glu Trp Gly Ala Phe Gly Asp Asp Gly Thr Leu Asn Asp
260 265 270

Ile Arg Thr Glu Phe Asp Arg Glu Ile Asp Met Gly Ser Leu Asn Pro
275 280 285

Gly Lys Gln Leu Phe Glu Lys Met Ile Ser Gly Met Tyr Met Gly Glu
290 295 300

Leu Val Arg Leu Ile Leu Val Lys Met Ala Lys Ala Glu Leu Leu Phe
305 310 315 320

Gln Gly Lys Leu Ser Pro Glu Leu Leu Thr Thr Gly Ser Phe Glu Thr
325 330 335

Lys Asp Val Ser Asp Ile Glu Glu Asp Lys Asp Gly Ile Glu Lys Ala
340 345 350

Tyr Gln Ile Leu Met Arg Leu Gly Leu Asn Pro Leu Gln Glu Asp Cys
355 360 365

Val Ala Thr His Arg Ile Cys Gln Ile Val Ser Thr Arg Ser Ala Ser
370 375 380

Leu Cys Ala Ala Thr Leu Ala Ala Val Leu Trp Arg Ile Lys Glu Asn
385 390 395 400

Lys Gly Glu Glu Arg Leu Arg Ser Thr Ile Gly Val Asp Gly Ser Val
405 410 415

Tyr Lys Lys His Pro His Phe Ala Lys Arg Leu His Lys Ala Val Arg
420 425 430

Arg Leu Val Pro Asp Cys Asp Val Arg Phe Leu Arg Ser Glu Asp Gly
435 440 445

Ser Gly Lys Gly Ala Ala Met Val Thr Ala Val Ala Tyr Arg Leu Ala
450 455 460

Asp Gln His Arg Ala Arg Gln Lys Thr Leu Glu Ser Leu Lys Leu Ser
465 470 475 480

His Glu Gln Leu Leu Glu Val Lys Arg Arg Met Lys Val Glu Met Glu
485 490 495

Gln Gly Leu Ser Lys Glu Thr His Ala Val Ala Pro Val Lys Met Leu
500 505 510

Pro Thr Tyr Val Cys Ala Thr Pro Asp Gly Thr Glu Lys Gly Asp Phe
515 520 525

Leu Ala Leu Asp Leu Gly Gly Thr Asn Phe Arg Val Leu Leu Val Arg
530 535 540

Val Arg Asn Gly Lys Arg Arg Gly Val Glu Met His Asn Lys Ile Tyr
545 550 555 560

Ser Ile Pro Gln Glu Val Met His Gly Thr Gly Glu Glu Leu Phe Asp
565 570 575

His Ile Val Gln Cys Ile Ala Asp Phe Leu Glu Tyr Met Gly Met Lys
580 585 590

Gly Val Ser Leu Pro Leu Gly Phe Thr Phe Ser Phe Pro Cys Gln Gln
595 600 605

Asn Ser Leu Asp Gln Ser Ile Leu Leu Lys Trp Thr Lys Gly Phe Lys
610 615 620

Ala Ser Gly Cys Glu Gly Glu Asp Val Val Thr Leu Leu Lys Glu Ala
625 630 635 640

Ile His Arg Arg Glu Glu Phe Asp Leu Asp Val Val Ala Val Val Asn
645 650 655

Asp Thr Val Gly Thr Met Met Thr Cys Gly Tyr Glu Asp Pro His Cys
660 665 670

Glu Val Gly Leu Ile Val Gly Thr Gly Ser Asn Ala Cys Tyr Met Glu
675 680 685

```
Glu Met Arg Asn Val Glu Leu Val Asp Gly Glu Glu Gly Arg Met Cys
    690                 695                 700

Val Asn Met Glu Trp Gly Ala Phe Gly Asp Asn Gly Cys Leu Asp Asp
705                 710                 715                 720

Leu Arg Thr Val Phe Asp Val Ala Val Asp Glu Leu Ser Leu Asn Pro
                725                 730                 735

Gly Lys Gln Arg Phe Glu Lys Met Ile Ser Gly Met Tyr Leu Gly Glu
            740                 745                 750

Ile Val Arg Asn Ile Leu Ile Asp Phe Thr Lys Arg Gly Leu Leu Phe
        755                 760                 765

Arg Gly Arg Ile Ser Glu Arg Leu Lys Thr Arg Gly Ile Ser Glu Thr
    770                 775                 780

Lys Phe Leu Ser Gln Ile Glu Ser Asp Cys Leu Ala Leu Leu Gln Val
785                 790                 795                 800

Arg Ala Ile Leu Arg His Leu Gly Leu Glu Ser Thr Cys Asp Asp Ser
                805                 810                 815

Ile Ile Val Lys Glu Val Cys Thr Val Val Ala Arg Arg Ala Ala Gln
            820                 825                 830

Leu Cys Gly Ala Gly Met Ala Ala Val Val Asp Lys Ile Arg Glu Asn
        835                 840                 845

Arg Gly Leu Asp Asn Pro Lys Val Thr Val Gly Val Asp Gly Thr Leu
    850                 855                 860

Tyr Lys Leu His Pro His Phe Ala Lys Val Met His Glu Thr Val Arg
865                 870                 875                 880

Asp Leu Ala Pro Lys Cys Asp Val Ser Phe Leu Glu Ser Glu Asp Gly
                885                 890                 895
```

Ser Gly Lys Gly Ala Ala Leu Ile Thr Ala Val Ala Cys Arg Ile Arg
900 905 910

Glu Ala Gly Gln Arg
915

We claim:

1. An isolated hexokinase II DNA fragment capable of regulating transcription of a downstream open reading frame, wherein the fragment comprises at least one of the response elements identified in Figure 11.
2. The fragment of claim 1 which is responsive to hypoxia.
3. The fragment of claim 1 which is responsive to p53.
4. The fragment of claim 1 which is responsive to glucose.
5. The fragment of claim 1 which is responsive to insulin.
6. The fragment of claim 1 which is responsive to AP-1.
7. The fragment of claim 1 which is responsive to AP-2.
8. The fragment of claim 1 which is responsive to ATF/CRE.
9. The fragment of claim 1 which has the sequence shown in SEQ ID NO:1 (Figure 5).
10. The fragment of claim 1 which is covalently joined to a reporter gene.
11. The fragment of claim 1 which is covalently joined to a toxic gene.
12. A method of screening for potential drugs which affect regulated transcription of tumor hexokinase II, the method comprising the steps of:

contacting a test substance with the reporter gene fusion of claim 10; and

measuring transcription of the reporter gene in the presence of the test substance; wherein a potential drug is identified when a test substance increases or decreases the transcription of the reporter gene.

13. The method of claim 12 wherein the transcription is performed in the presence of a transcription factor which binds to the response element.
14. A method of treating cells which overexpress hexokinase II comprising the step of:

administering the gene fusion of claim 11 to cells which overexpress hexokinase II, whereby the toxic gene is expressed in the cells.

15. An isolated nucleic acid probe comprising at least 15 contiguous nucleotides selected from the sequence of SEQ ID NO: 1.

16. The isolated nucleic acid probe of claim 15 wherein the probe is selected from the nucleotides number -4369 to -1158.

17. The isolated nucleic acid probe of claim 15 which is detectably labeled.

18. A method for diagnosing tumors which overexpress hexokinase, comprising the steps of:

determining copy number of a hexokinase II gene in a tissue sample suspected of being neoplastic; wherein a determined copy number of greater than two indicates neoplasia.

19. The method of claim 18 wherein a copy number of greater than ten indicates neoplasia.

20. The method of claim 18 wherein copy number is determined by hybridization to a nucleic acid probe comprising at least 15 contiguous nucleotides seleted from the sequence of SEQ ID NO:1 wherein the probe is detectably labeled.

21. The method of claim 18 wherein copy number is determined by comparing hybridization of a hexokinase II gene probe with a normal tissue to hybridization of the probe with the tissue sample suspected of being neoplastic.

22. The method of claim 18 wherein copy number is determined by a quantitative polymerase chain reaction.

23. The method of claim 18 wherein copy number is determined by fluorescence *in situ* hybridization (FISH) analysis.

24. A method for diagnosing neoplastic tissues, comprising the step of:

determining whether cells in a tissue sample suspected of being neoplastic contain a hexokinase II gene which is unmethylated, an unmethylated hexokinase II gene indicating neoplasia.

25. The method of claim 24 wherein the step of determining whether the hexokinase II gene is unmethylated is performed by use of restriction endonucleases which are methylation sensitive.

26. A vector for expression of a desired protein in a mammalian cell, comprising:

an isolated hexokinase II DNA fragment according to claim 1.

27. The vector of claim 26 further comprising:

a gene encoding the desired protein, wherein the gene is covalently linked to the DNA fragment and the DNA fragment regulates the expression of the desired gene in a mammalian cell.

28. The vector of claim 27 wherein the DNA fragment is upstream from the desired gene.

29. A method for increasing glycolysis in cells, comprising the step of:

introducing into cells an unmethylated DNA molecule comprising:

a hexokinase II DNA fragment capable of regulating transcription of a downstream open reading frame, wherein the fragment comprises at least one of the response elements identified in Figure 11; and

a nucleic acid encoding a hexokinase II, wherein the hexokinase II DNA fragment is covalently and operatively linked to the nucleic acid encoding a hexokinase II.

30. The method of claim 29 wherein the cells are liver cells.

31. The method of claim 29 wherein the cells are muscle cells.

32. The method of claim 29 wherein the cells are adipose cells.

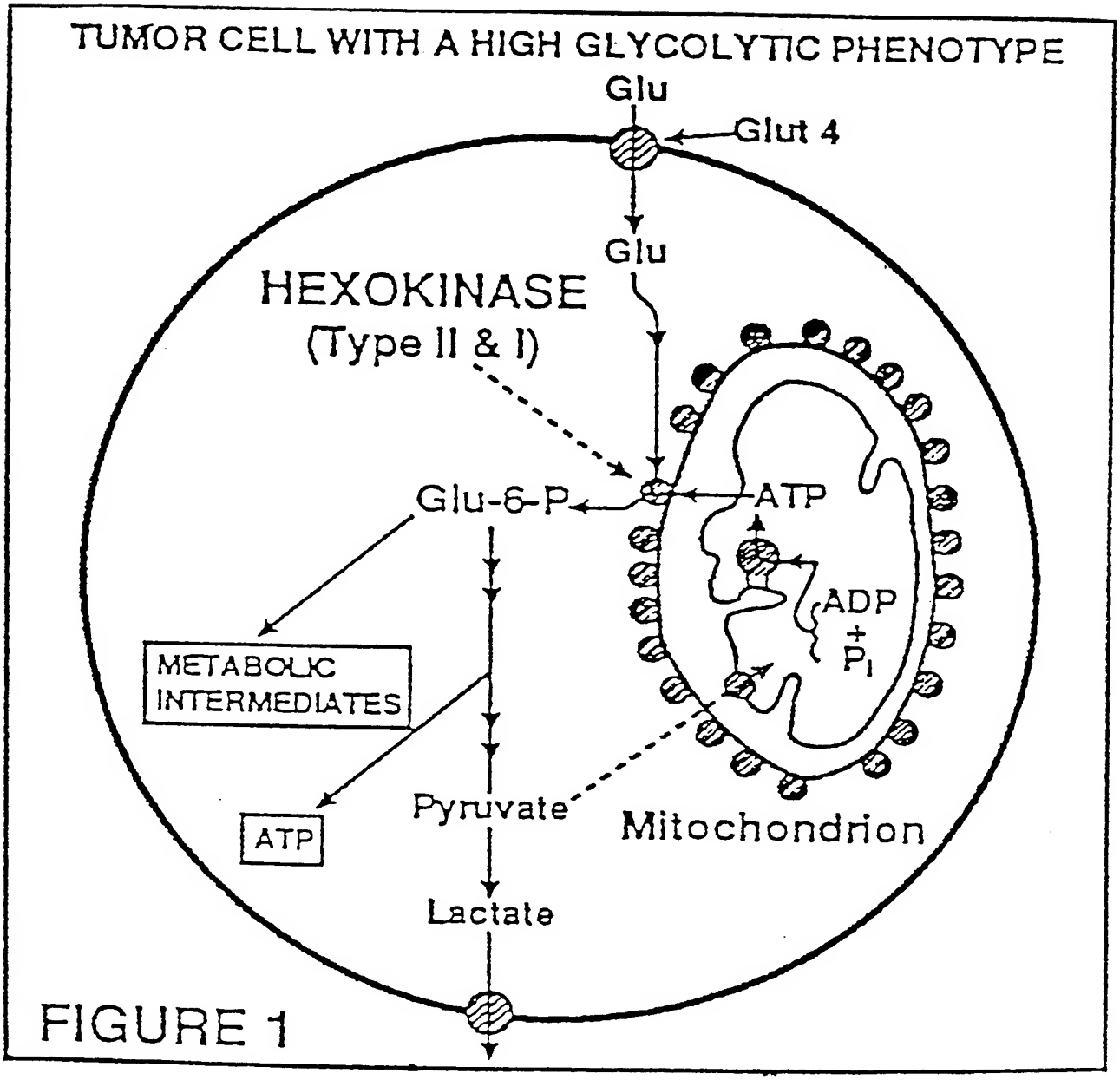
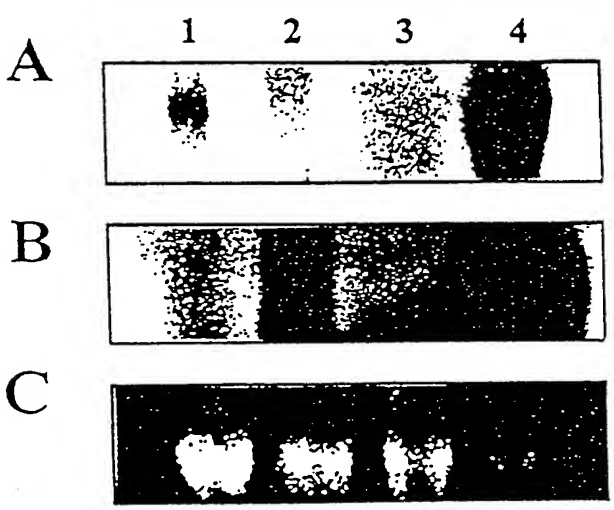

TUMOR CELL WITH A HIGH GLYCOLYTIC PHENOTYPE
Glu
Glut 4
Glu
HEXOKINASE
(Type II & I)
Glu-6-P
ATP
ADP
+
Pi
METABOLIC
INTERMEDIATES
ATP
Pyruvate
Mitochondrion
Lactate
FIGURE 1

FIGURE 1

FIGURE 2



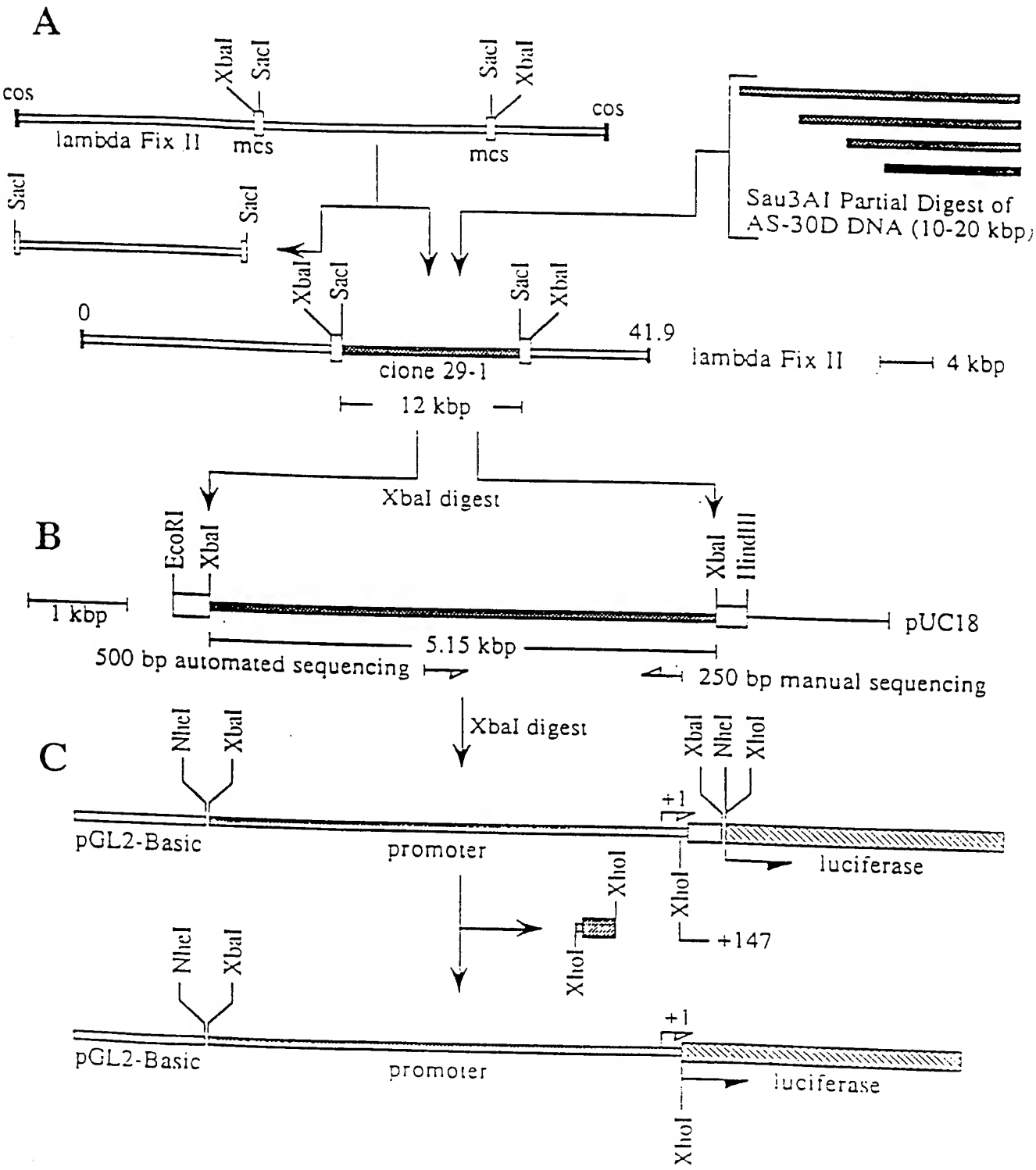

A
cos
XbaI
SacI
SacI
XbaI
cos
lambda Fix II
mcs
mcs
Sau3AI Partial Digest of
AS-30D DNA (10-20 kbp)
SacI
SacI
XbaI
SacI
SacI
XbaI
0
41.9
clone 29-1
lambda Fix II
4 kbp
12 kbp
XbaI digest
B
EcoRI
XbaI
XbaI
HindIII
1 kbp
pUC18
5.15 kbp
500 bp automated sequencing
250 bp manual sequencing
XbaI digest
C
NheI
XbaI
XbaI
NheI
XhoI
+1
pGL2-Basic
promoter
luciferase
XhoI
XhoI
+147
NheI
XbaI
+1
pGL2-Basic
promoter
luciferase
XhoI

FIGURE 3

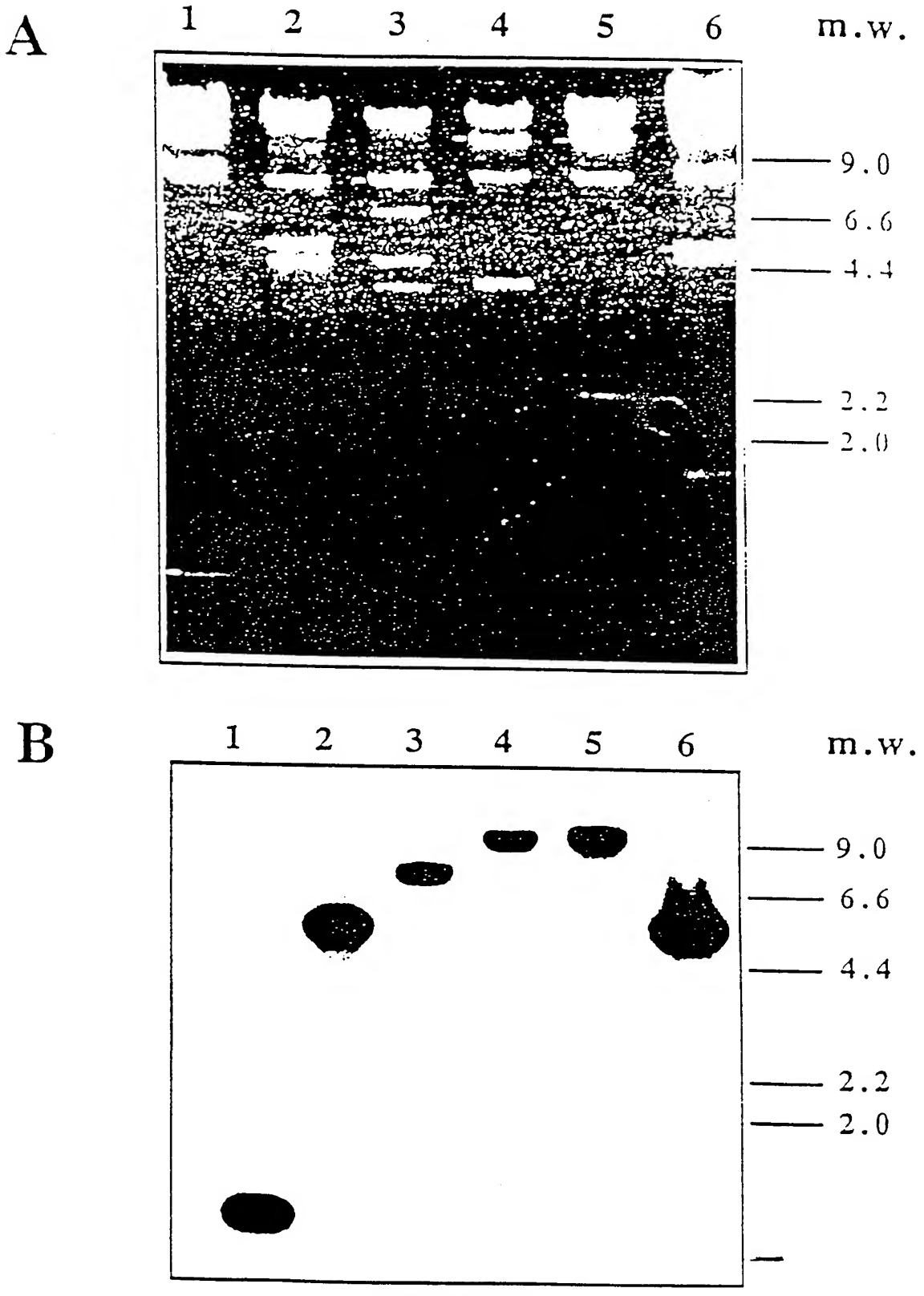


FIGURE 4

```
-4369  TCTAGAGCTCGCGGCCGCGAGCTCTAATACGACTCACTATAGGGCGTCGA  -4320

-4319  CTCGATCCAACTGGCCTAGAACTCACAGCCATCCTCTTGCCTCTACCTAT  -4270
             Myb

-4269  GGAGTGTGGAGATTAAAGGCATGTTCTACCATGTCTTAATTTTAAAATAC  -4220
                           p53

-4219  CTATGGAGTGTGGAGATTAAAGGCATGTTCTACCATGTCTTAATTTTAAA  -4170
                           p53

-4169  ATAGGAATATTTGTGGATTGAGGTCTTGAGCAAAATAAGATTTTTCCCAA  -4120
          HNF-5,Ap-3,c/ebp          IRE

-4119  GAGAGTTTCCTGAAGCCTAAGTAGACTCAGGTCCTTCTCATGCAGGGCCA  -4070

-4069  ATCTAGGGCCAGGAGCAGGACCAACTGGTGTGAAATCAGAAAGATGGTAC  -4020
                           Myb

-4019  TCATAGCTATTAGTCCATCTCTGGTTGACACTCCCAGACTCCCCTACATC  -3970

-3969  TCAAGACACAGACATACGTGGCTTTTTATGAATCCATTTTTCTGGTCTGT  -3920

-3919  ATTATTTGTTCTGTGTGTTAATTTTATGTCTTAAACTAAACAGAAATCCT  -3870
          HNF-5

-3869  TTAAGGAAAGAACACCCCGCCCCTCTCCTGTGTGTGTGTGTGTGTGTGTG  -3820
           Pea-3, SRE, Ap-2

-3819  TGTGTGTGTACACGTCTGTGTGAGTATCTCGCACCCTGTAAAGGGCTTAA  -3770

-3769  TAAACACGTGCTGATTGATTCCCCTCCTGGAATGTGAATGTGGACTGCCA  -3720
          GlRE, Myc                                c/ebp

-3719  ATCTGCCAGTCTACAATGTGTGTGCCTGTATGTGCTCATGGGGAGAGAGA  -3670

-3669  GGGAGAAATAAAATAGACTCTAAGGAAGAATCTTGAGGACAGGAAAGTCA  -3620
                           Pea-3              Pea-3

-3619  GAGCTACACACCTCACTTTTGAGTGGGTAGCTGTCCCCTGATTTGACACA  -3570

-3569  TACAGATGGGTTAGGGGATATCACTGTACTCACTCCAGCCACCTCCCAGG  -3520

-3519  GTTACTGGGAACTCTGTGAGAGATCATCCCATAAAGTACCCTGTGAACAT  -3470

-3469  GAGTTAGTCCTCATAAAGTGGGACCAGAAAAGAGAATGGAGAATGGAGCT  -3420
          Ap-1
```

FIGURE 5 A

-3419 GAAGTGTGTGTGCAAGTAAGTGTGTGTGAGATCCAGCTAATTGGACTCAG -3370

-3369 CTGATGGAGTGCTTGCCTAGCACGCATGAATCCTCATGTTTGCCTCTGAT -3320
HNF-5

-3319 CGCAAGACCTGAAAAAAAAAAAAAATAGGCGAGGTAGACAGTGCCTGTAA -3270
Sp-1

-3269 CCTCAGCGCTGAGGAAGTGGAGGCCGGAGGATGGGAAGCTCAAGACTGTC -3220
Pea-3

-3219 CTTGGTTGCATGTTTAGTTAGAGGCCATCTTGGGCCACATGATCCTGTCC -3170

-3169 CAAAATAAACAAAGGAATACAATTAGTCCATAGGGAGGAGATCATAGTTG -3120

-3119 ACCTGACCCCACTGATTTTGATCTTAGTTGTCTGAGGGAAATTATTTTAT -3070

-3069 ATACTGATTTAACTATCGGTTTTTTAAGTGTCTCAAAATGTTTTTATTTC -3020

-3019 ATGTACACCCTTATTGGGTGTGCATATGCATGTAGGTACACACATGCCAT -2970

-2969 GGTAAGTGTGTAAAGGTCAGAAGTCAATTTTCTTGAGTTGATTCTCTCCT -2920
PPAR

-2919 GTGACCACATGGGTCCTAGGGTTCACCTCAAGTAGTCGGGGTTCAGCGAC -2870

-2869 AAGCGCCTTTACCCACCGAGCCACCTTGCCAGCACCCGAAGTGTTTCCAG -2820

-2819 AAAGGTCTTTTTTTTTTTCTCTTTGCTGCTTACTTTTAACCTATGCCATC -2770

-2769 AATTCTGCCTCAGACTTCTGAACACCTAAAGCCTTAATCAGCCTCTGTGC -2720
Ap-1

-2719 CTCACCCTTGTCTCACTCCAGCCTTTATCTTATCTGGGAGTTCCTGTCTC -2670

-2669 TTCTCCTTCAGGCCGGGTCCTTTCCTCCCATTCATGTGGAGAGCAGCTTT -2620

-2619 TGTCCTACAAAAGCTTTAAGCATCTCAGAGTCTGTGTCAGAAAGAGAGGA -2570

-2569 GCTGGCTTATGAGGCTGTTGCAATTGGGTGAAAGACACTGGTGAACTGTG -2520
c/ebp,NF-IL6

-2519 AGGCAGACCAATGGGAAGGGTTTGAGAACTAATATAGAAAATGAAAGTCT -2470

-2469 CTCCTTTGTGTCGTATAATCATATGTGACATCACTAAATCATCTACTAAC -2420

FIGURE 5 B

```
-2419  TTACACAATAAATACCTACATGGTGCCTACCATGTGATAGAGCGCCCCCA  -2370
-2369  CACGAGGTACTGCAGATAGAAGGGAATGATATAGACGCAGATGCTTATTC  -2320
                                                  Ap-1
-2319  AGACAGGTAGGACAGAATGGATATAACACTTAGAAAAGGACCCGGGTGTG  -2270
                                                  I----------
-2269  GTGGCACGGTGGCACATACCTTAGATCCCACCACCCGGGTGTTGGGGCTG  -2220
                                          I----------  Sp-1
-2219  AGGCAGATGACTTTTGTTTGTTTTTTTGTTTTCAGTTTTGCTTTTTTTCA  -2170
                      HNF-5
-2169  AGACAGGGTCTCTCCGTGCAGCACTGCCTGTCCTGGAACTCGCTTTGCTG  -2120
-2119  GCCACGTTTGTGGCCTTGAACTCACAAAGTGCTGGTGCCTGGCTGTAAAA  -2070
-2069  TTAATTTCTCTCTCCCTCTCTCCCCCCCTCCCCCACCTCTCTCGCTACTT  -2020
                                  Ap-2
-2019  GCTTGGTAGACCAGACTGGCTTCGAACTCAGAGATTTGCCTGCGTTTGCC  -1970
-1969  GCCCAAGGGCTGTGATTAAAGGTATGTGCCACCATGTCCAGCCTTAAAAA  -1920
         Ap-2      Ap-1
-1919  TTACTTCTAATAGTCATTCTTAGGAGTTTGGATTTTATTTGAAGATAAGA  -1870
-1869  AAACAATAATGGTTTTAAGACTCTTCCCCCCCAAAAAGAGAGTTTGGTAT  -1820
-1819  ATATCTATCAATCAATCTAATCTTATCTCCTGCCTGCCTGCGTATCTATC  -1770
                                                  II--------
-1769  TATCTATCTATCTATCTATCTATCTATCTATCCATCCATCCAAGGT      -1720
       --------  II--------------------
-1719  CTCATGCTTACCAAGTTGGGCTTGAACTCCTGACTCTTCTGTCTCCACCT  -1670
-1669  CTGGAGCACTGGGACTACAAATTTGTGCCACCCCACAAAGCACTGGCTTG  -1620
-1619  TATTTTAAACAAGTCTCTTTAGCTCTTGAGTAAGAGGGTTCATGGTGGTC  -1570
                 p53              NF-IL6,Ap-1
-1569  AAACTAGAGGTAGCTAAAAATGGCAGCTAAGTGACATTACACGGACTCGG  -1520
-1519  GTGGAGTCATCGATGGCCTGGCCATGAGGGTCTGGCCTTTTTGATTTGCA  -1470
                        Ap-2
-1469  GTTAGACTAACTGCTCCCCGATGGAGTGGATAGTTGTAAGAGCAGGTGGC  -1420
       Myb,Mbf-1,Mep-1,Mtf-1      c/ebp
```

FIGURE 5 C

```
-1419  AGGAAGACACCATGGATGGTGATGTCATTTGTGGAGACAACTGGTAAAGG  -1370
       Pea-3                                    Myb  Pea-3

-1369  AAAAAAAAAACCTAAGAAGTTCAGCTTAGTACATGTTAAGTGTGAGGTGC  -1320

-1319  TAGTCACCTCTGCTGGGATGACAAACCATAGCTGGCTAAGAAAGTTAGAA  -1270
                                    SRE

-1269  GCCCTGGGGGAAAGTTCTTGCTTTGTAGATTAGGTTTGTGAGTACCTGTG  -1220
         Ap-2

-1219  TGCAGATGGTGTAACTGCTGTCGACAGTGCTGGGAGGAATCGCCCATCGA  -1170
                   Myb

-1169  GGGAATAGATGAACCACACCAGAGAACATGGTAGAAGCGGCCCAGCAGAG  -1120
                                 SRE

-1119  CAACGTGGGCTGGGGTGTACTTCAGTCGGCAGAGTGCTTTGATCTCCAGT  -1070
           Sp-1,Ap-2,SRE

-1069  AGTGGCCCATGCCACTCCAGAGGTGGGGAGAGAGCTTGGGGAGCGAGACT  -1020
             c/ebp

-1019  GTTTGGAAATGGTAGGCCCTGTCTTCTTTCCATGTAACTTTCCAACTCCC  -970
         NF-IL6

 -969  AGGTTTCCCATTCTCCACCAGCAACAACCTCATGCCATTTGAGGTGCTAC  -920

 -919  TTCAATATCGCTGGCGTCTACTCATCTATGTGAACTTAAGAGTCTACTCA  -870
                   III----------------  III---------

 -869  TCTATGTGACTAAGAGTCTGGTGTCAGGCGTGAGCTGAGGTAGAGGTGGG  -820
       ---------Ap-1

 -819  CTCTTCTCAGCCTCTATAAACCAATTCACACCACTTGAGCCAAGCAGTTA  -770

 -769  CACATGCACTTTCTCCTCCCGCCTATCAGTCCTAGCTCCTGACAAGGTTT  -720

 -719  CTCTCCAGCCTTTTACTTTCCTGGCTTCAAGAAAGGCGGGATAATATACC  -670

 -669  AGGGTGGGGAGATTGCATTTCAGAGTGAGACGTGTTCTGTCTTCACCTAC  -620
         Ap-2      c/ebp                              c/ebp

 -619  CACTTGTTGGCGATGTGACCTTGGGCAAAGCTCATTAACAGCACAGTGCC  -570
             F-ACT1

 -569  TAGTTCCCTAATTTGTAAAACATATGCTATAGGTGTGACGATTACGAAGG  -520
                                          cre          -

 -519  GCTGACTTTTGTAATGGCTTTGCTTCAGGGATCTGCAGACTCGTTGAGCC  -470
             NF-IL6
```

FIGURE 5 D

```
-469  ACAATTAGGATGAGAATCAAGGTGCTTCAGACTTGTGACAGGGCACTGGC   -420

-419  GGCCCCTCACATGATCCTCAGATACCAGATTGTGGCGTGTGCTGCTAGGA   -370
                                      SRE

-369  TCACTTGTCTTTCCAGTCTCCCAACATCTCTTGGGTCCGTGATCACGCGC   -320

-319  CCCCCACCCGAAGCCCAGCCTGACGCGGCGGTGGCTCATGCGCCCTGGAG   -270
        Ap-2              cre/atf-1

-269  TCCCGGGCTCTAGCCACGGAACACACGTCCCAACTCTGGCGCCCGGCTCC   -220
            c/ebp

-219  GCCCCTAGCCTCGGGCGCGTCTCTCCCGCCGCCTGCTTGGGTGCTGGAGC   -170

-169  AGCCGCGCCCGCGGGCTCTGGGCGCTGATTGGCTGTGGACTGCGGGCGGG   -120

-119  CAGCCGGAGAGCGTACACACCCTCTTCCCGCAGCCAATGAGCGCGCCCAC   -70
                                        -85          atf-1

 -69  GTCACTGTCTTGGGCGGCCCAAAGAGCCGGCAGCCCCTCAATAAGCCACA   -20
          Sp-1              Ap-2       -30

 -19  TTGTTGCACCAACTCCAGTGCTAGAGTCTCAGGACACCACAGGCTACACG    31
                        +1

  32  GAGTTATCCCGCTTAGGAGACCCGAAGGCAGGAGCATCACTCCAGTGACT    81

  82  CTGATAAGGTGCGATCGCCCGAGAGGAACAGAACTGTCATTTTTGCGAAG   131

 132  TTGAGCCTTACGGATCCCGTGGGCGAAGTTAGCGACGGGACGCTGAGCAA   181

 182  CTAGACCGGACGGCAGGAGTGAGACTTAGGTGCCTTCTAGTAGTTGTGAC   231

 232  TTAAAAAAAAAAAAAAAAGGAAAAGAAAAAAGGAGGAAAACCTGTTTCTG   281

 282  GAAACGCGAGGCCCTCAGCTGGTGAGCCATCGTGGTTAAGCTTCTTTGTG   331

 332  TGGCTCCTGGAGTCTCCGATCCCAGCCGGACACCCGGGCCTGGTTTCAAA   381

 382  GCGGTCGAACTGCTCTGCCCGCTCCACCGGTAGCGCTCGAGCCTCGGTTT   431

 432  CTCTACTCGACCCCGACTCGCCGCAGCAGGATGATCGCCTCGCATATGAT   481
                                    MetIleAlaSerHisMetIl
```

FIGURE 5 E

```
482  CGCCTGCTTATTCACGGAGCTCAACCAAAACCAAGTGCAGAAGGTAAGTC  531
     eAlaCysLeuPheThrGluLeuAsnGlnAsnGlnValGlnLys

532  GGCACGGGCGGGAGCTGCTGGCTCGCTTCGGACCAAGTTGCGTGCTCTCC  581
       GlRE

582  GGGAATCTGGAGCACGCAGAGGACCTGCTTCCTCCTCCGGGGCTGGGGAC  631
                                              Ap-2

632  GTGGAACCAGTCTGAGTAGCTGGGAAAGTCCTGAGCGCCAGAAACCACGT  681

682  CTGCTAGGCACCCTCGTGGCCCGGCCGCGCATCACCGATACTCCCACTTT  731

732  CCCGGGATCCGCGAGCATCCTCCCCACCCTTAAAGCCCCTAATTTCTAGA  781
```

FIGURE 5F

```
-17   CGACTCGCCGCAGCAGGATGATCGCCTCGCATATGATCGCCTGCTTATTCACGGAGCTC      42
                        MetIleAlaSerHisMetIleAlaCysLeuPheThrGluLeu

 43   AACCAAAACCAAGTGCAGAAGGTTGACCAATTTCTCTACCACATGCGTCTCTCAGATGAG     102
      AsnGlnAsnGlnValGlnLysValAspGlnPheLeuTyrHisMetArgLeuSerAspGlu

103   ACCCTTCTGGAGATTTCTAGGCGGTTCCGGAAGGAGATGGAGAAAGGGCTAGGAGCTACC     162
      ThrLeuLeuGluIleSerArgArgPheArgLysGluMetGluLysGlyLeuGlyAlaThr

163   ACGCACCCTACAGCAGCTGTGAAAATGTTGCCTACCTTTGTGAGGTCAACTCCGGATGGG     222
      ThrHisProThrAlaAlaValLysMetLeuProThrPheValArgSerThrProAspGly

223   ACAGAACATGGGGAGTTCCTGGCTCTGGATCTTGGAGGAACCAACTTCCGTGTGCTCCGA     282
      ThrGluHisGlyGluPheLeuAlaLeuAspLeuGlyGlyThrAsnPheArgValLeuArg

283   GTAAGGGTGACGGACAATGGCCTCCAGAGAGTGGAGATGGAGAACCAGATCTACGCCATC     342
      ValArgValThrAspAsnGlyLeuGlnArgValGluMetGluAsnGlnIleTyrAlaIle

343   CTTGAGGACATCATGCGGGGCAGTGGAACCCAGCTGTTTGACCACATCGCCGAATGCCTG     402
      LeuGluAspIleMetArgGlySerGlyThrGlnLeuPheAspHisIleAlaGluCysLeu

403   GCCAACTTCATGGACAAGCTACAAATCAAAGAGAAGAAGCTCCCTCTGGGTTTCACCTTC     462
      AlaAsnPheMetAspLysLeuGlnIleLysGluLysLysLeuProLeuGlyPheThrPhe

463   TCGTTCCCCTGCCACCAGACAAAACTGGATGAGAGTTTTTTGGTCTCGTGGACTAAGGGG     522
      SerPheProCysHisGlnThrLysLeuAspGluSerPheLeuValSerTrpThrLysGly

523   TTCAAGTCCAGTGGCGTGGAAGGCAGAGATGTGGTGGACCTGATCCGGAAGGTTATCCAG     582
      PheLysSerSerGlyValGluGlyArgAspValValAspLeuIleArgLysValIleGln

583   CGCAGAGGGGACTTTGACATTGACATTGTGGCCGTGGTGAATGACACAGTTGGGACCATG     642
      ArgArgGlyAspPheAspIleAspIleValAlaValValAsnAspThrValGlyThrMet

643   ATGACTTGTGGCTATGATGATCAGAACTGCGAGATTGGTCTCATTGTGGGCACTGGCAGC     702
      MetThrCysGlyTyrAspAspGlnAsnCysGluIleGlyLeuIleValGlyThrGlySer

703   AACGCCTGCTACATGGAGGAAATGCGTCATATTGACATGGTGGAGGGAGATGAGGGGCGC     762
      AsnAlaCysTyrMetGluGluMetArgHisIleAspMetValGluGlyAspGluGlyArg

763   ATGTGCATCAACATGGAGTGGGGAGCCTTTGGGGACGACGGTACACTCAATGACATCCGA     822
      MetCysIleAsnMetGluTrpGlyAlaPheGlyAspAspGlyThrLeuAsnAspIleArg

823   ACCGAGTTTGACCGAGAGATCGACATGGGCTCGCTGAACCCTGGGAAGCAGCTGTTTGAG     882
      ThrGluPheAspArgGluIleAspMetGlySerLeuAsnProGlyLysGlnLeuPheGlu

883   AAGATGATTAGCGGGATGTACATGGGGGAGCTGGTCAGGCTCATCCTGGTGAAGATGGCC     942
      LysMetIleSerGlyMetTyrMetGlyGluLeuValArgLeuIleLeuValLysMetAla
```

FIGURE 6A

```
943  AAGGCAGAGCTGTTGTTCCAAGGGAAACTCAGCCCAGAACTCCTTACCACTGGCTCCTTC  1002
     LysAlaGluLeuLeuPheGlnGlyLysLeuSerProGluLeuLeuThrThrGlySerPhe

1003 GAGACCAAAGATGTCTCGGATATTGAAGAGGATAAGGATGGAATCGAGAAGGCCTACCAA  1062
     GluThrLysAspValSerAspIleGluGluAspLysAspGlyIleGluLysAlaTyrGln

1063 ATCCTGATGCGCCTGGGTCTGAATCCATTGCAGGAGGATTGTGTGGCCACGCACCGAATC  1122
     IleLeuMetArgLeuGlyLeuAsnProLeuGlnGluAspCysValAlaThrHisArgIle

1123 TGCCAGATTGTGTCCACGCGCTCGGCCAGTCTGTGCGCAGCCACCCTGGCCGCGGTGCTG  1182
     CysGlnIleValSerThrArgSerAlaSerLeuCysAlaAlaThrLeuAlaAlaValLeu

1183 TGGCGAATCAAAGAGAACAAGGGCGAGGAGCGACTTCGCTCCACCATCGGTGTCGATGGC  1242
     TrpArgIleLysGluAsnLysGlyGluGluArgLeuArgSerThrIleGlyValAspGly

1243 TCCGTCTACAAGAAACATCCCCATTTTGCCAAGCGTCTCCATAAGGCAGTGAGGAGGCTG  1302
     SerValTyrLysLysHisProHisPheAlaLysArgLeuHisLysAlaValArgArgLeu

1303 GTGCCCGACTGTGATGTCCGCTTCCTCCGCTCTGAGGATGGCAGCGGCAAGGGGGCTGCT  1362
     ValProAspCysAspValArgPheLeuArgSerGluAspGlySerGlyLysGlyAlaAla

1363 ATGGTGACGGCGGTGGCTTACCGTCTGGCTGACCAACACCGGGCCCGCCAGAAGACCCTG  1422
     MetValThrAlaValAlaTyrArgLeuAlaAspGlnHisArgAlaArgGlnLysThrLeu

1423 GAGTCTCTGAAGCTGAGCCACGAGCAGCTTCTGGAGGTTAAGAGAAGAATGAAGGTGGAA  1482
     GluSerLeuLysLeuSerHisGluGlnLeuLeuGluValLysArgArgMetLysValGlu

1483 ATGGAGCAGGGTCTGAGCAAGGAGACGCATGCGGTCGCCCCTGTGAAGATGCTGCCCACT  1542
     MetGluGlnGlyLeuSerLysGluThrHisAlaValAlaProValLysMetLeuProThr

1543 TACGTGTGTGCCACTCCAGATGGCACAGAGAAAGGAGACTTCTTGGCCTTGGATCTTGGA  1602
     TyrValCysAlaThrProAspGlyThrGluLysGlyAspPheLeuAlaLeuAspLeuGly

1603 GGAACAAACTTCCGGGTCCTGCTGGTGCGTGTGCGTAATGGCAAGCGGAGGGGCGTGGAG  1662
     GlyThrAsnPheArgValLeuLeuValArgValArgAsnGlyLysArgArgGlyValGlu

1663 ATGCATAACAAGATCTACTCCATCCCACAGGAGGTTATGCATGGCACTGGGGAAGAGCTC  1722
     MetHisAsnLysIleTyrSerIleProGlnGluValMetHisGlyThrGlyGluGluLeu

1723 TTCGACCACATTGTCCAGTGCATTGCGGACTTCCTGGAGTACATGGGCATGAAGGGCGTG  1782
     PheAspHisIleValGlnCysIleAlaAspPheLeuGluTyrMetGlyMetLysGlyVal

1783 TCCCTGCCTTTGGGTTTCACATTCTCCTTCCCTTGCCAGCAGAACAGCCTAGACCAGAGC  1842
     SerLeuProLeuGlyPheThrPheSerPheProCysGlnGlnAsnSerLeuAspGlnSer

1843 ATCCTCCTCAAGTGGACAAAGGGATTCAAGGCATCTGGCTGCGAGGGTGAGGATGTGGTC  1902
     IleLeuLeuLysTrpThrLysGlyPheLysAlaSerGlyCysGluGlyGluAspValVal

1903 ACCTTGCTGAAGGAAGCGATTCACCGGCGAGAGGAGTTTGACCTGGATGTGGTTGCCGTG  1962
     ThrLeuLeuLysGluAlaIleHisArgArgGluGluPheAspLeuAspValValAlaVal
```

FIGURE 6B

```
1963  GTGAATGACACAGTTGGGACTATGATGACTTGTGGCTACGAAGACCCTCACTGTGAAGTT  2022
      ValAsnAspThrValGlyThrMetMetThrCysGlyTyrGluAspProHisCysGluVal

2023  GGCCTCATTGTTGGCACCGGAAGCAACGCCTGCTACATGGAAGAGATGCGTAATGTGGAG  2082
      GlyLeuIleValGlyThrGlySerAsnAlaCysTyrMetGluGluMetArgAsnValGlu

2083  CTGGTGGACGGAGAGGAGGGACGGATGTGTGTCAACATGGAGTGGGGAGCATTTGGGGAC  2142
      LeuValAspGlyGluGluGlyArgMetCysValAsnMetGluTrpGlyAlaPheGlyAsp

2143  AATGGCTGCCTGGATGACTTGCGGACCGTGTTTGATGTTGCTGTGGATGAGCTTTCTCTC  2202
      AsnGlyCysLeuAspAspLeuArgThrValPheAspValAlaValAspGluLeuSerLeu

2203  AACCCTGGCAAACAGAGGTTCGAGAAGATGATCAGCGGCATGTACTTGGGAGAGATTGTG  2262
      AsnProGlyLysGlnArgPheGluLysMetIleSerGlyMetTyrLeuGlyGluIleVal

2263  CGCAACATTCTCATCGATTTCACGAAGCGGGGGCTGCTCTTCCGAGGCCGCATCTCAGAG  2322
      ArgAsnIleLeuIleAspPheThrLysArgGlyLeuLeuPheArgGlyArgIleSerGlu

2323  CGCCTCAAGACAAGGGGAATCTCTGAAACTAAGTTCCTGTCTCAGATAGAGAGCGACTGC  2382
      ArgLeuLysThrArgGlyIleSerGluThrLysPheLeuSerGlnIleGluSerAspCys

2383  CTAGCCCTGCTACAGGTTCGTGCCATCCTGCGCCACCTAGGGCTGGAGAGCACGTGCGAT  2442
      LeuAlaLeuLeuGlnValArgAlaIleLeuArgHisLeuGlyLeuGluSerThrCysAsp

2443  GACAGCATCATCGTGAAGGAGGTGTGCACTGTGGTTGCCCGGCGCGCTGCACAGCTCTGT  2502
      AspSerIleIleValLysGluValCysThrValValAlaArgArgAlaAlaGlnLeuCys

2503  GGCGCAGGCATGGCCGCCGTAGTGGACAAGATAAGAGAGAACCGTGGGCTGGACAACCCC  2562
      GlyAlaGlyMetAlaAlaValValAspLysIleArgGluAsnArgGlyLeuAspAsnPro

2563  AAAGTGACAGTGGGCGTGGACGGGACTCTGTATAAGCTTCATCCTCACTTTGCCAAGGTC  2622
      LysValThrValGlyValAspGlyThrLeuTyrLysLeuHisProHisPheAlaLysVal

2623  ATGCATGAGACGGTGAGAGATCTGGCTCCGAAATGTGACGTGTCCTTCCTGGAATCCGAG  2682
      MetHisGluThrValArgAspLeuAlaProLysCysAspValSerPheLeuGluSerGlu

2683  GACGGCAGTGGGAAGGGAGCAGCTCTCATCACTGCCGTGGCCTGCCGCATCCGGGAGGCT  2742
      AspGlySerGlyLysGlyAlaAlaLeuIleThrAlaValAlaCysArgIleArgGluAla

2743  GGGCAGAGATAG
      GlyGlnArg *
```

FIGURE 6C

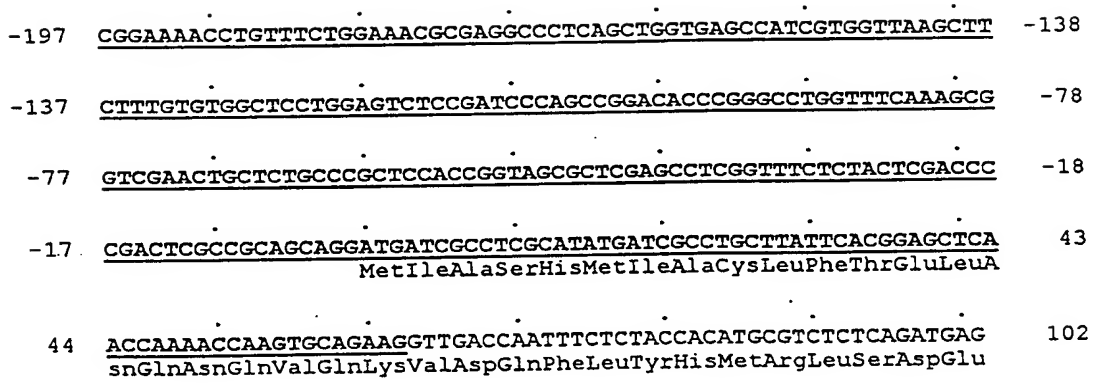

-197 CGGAAAACCTGTTTCTGGAAACGCGAGGCCCTCAGCTGGTGAGCCATCGTGGTTAAGCTT -138
-137 CTTTGTGTGGCTCCTGGAGTCTCCGATCCCAGCCGGACACCCGGGCCTGGTTTCAAAGCG -78
-77 GTCGAACTGCTCTGCCCGCTCCACCGGTAGCGCTCGAGCCTCGGTTTCTCTACTCGACCC -18
-17 CGACTCGCCGCAGCAGGATGATCGCCTCGCATATGATCGCCTGCTTATTCACGGAGCTCA 43
MetIleAlaSerHisMetIleAlaCysLeuPheThrGluLeuA
44 ACCAAAACCAAGTGCAGAAGGTTGACCAATTTCTCTACCACATGCGTCTCTCAGATGAG 102
snGlnAsnGlnValGlnLysValAspGlnPheLeuTyrHisMetArgLeuSerAspGlu

FIGURE 7

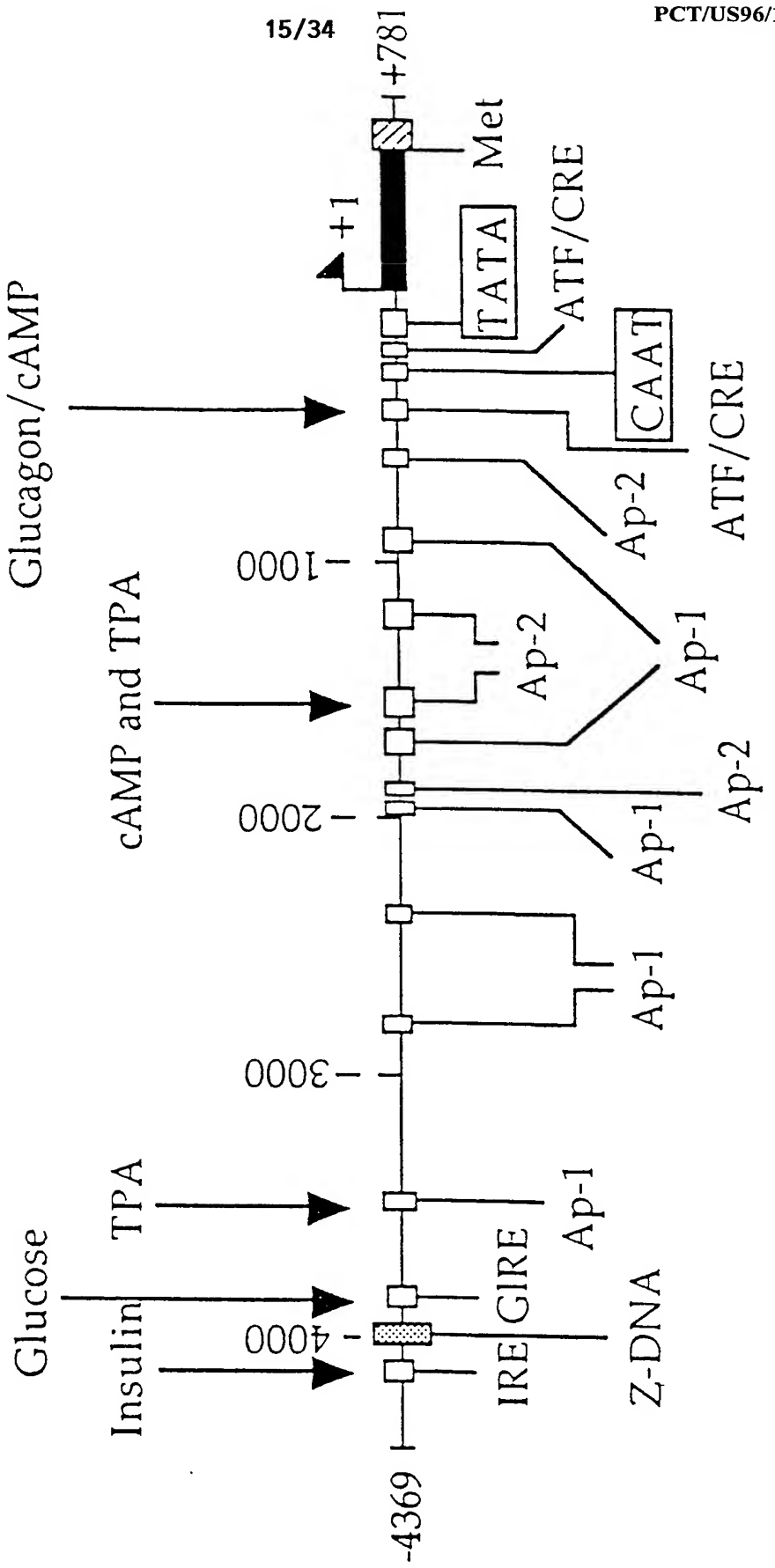

+781
Met
+1
TATA
ATF/CRE
CAAT
Glucagon/cAMP
-1000
Ap-2
ATF/CRE
cAMP and TPA
Ap-2
Ap-1
-2000
Ap-1
Ap-2
Ap-1
-3000
TPA
Ap-1
Glucose
Insulin
-4000
GIRE
IRE
Z-DNA
-4369

FIGURE 8

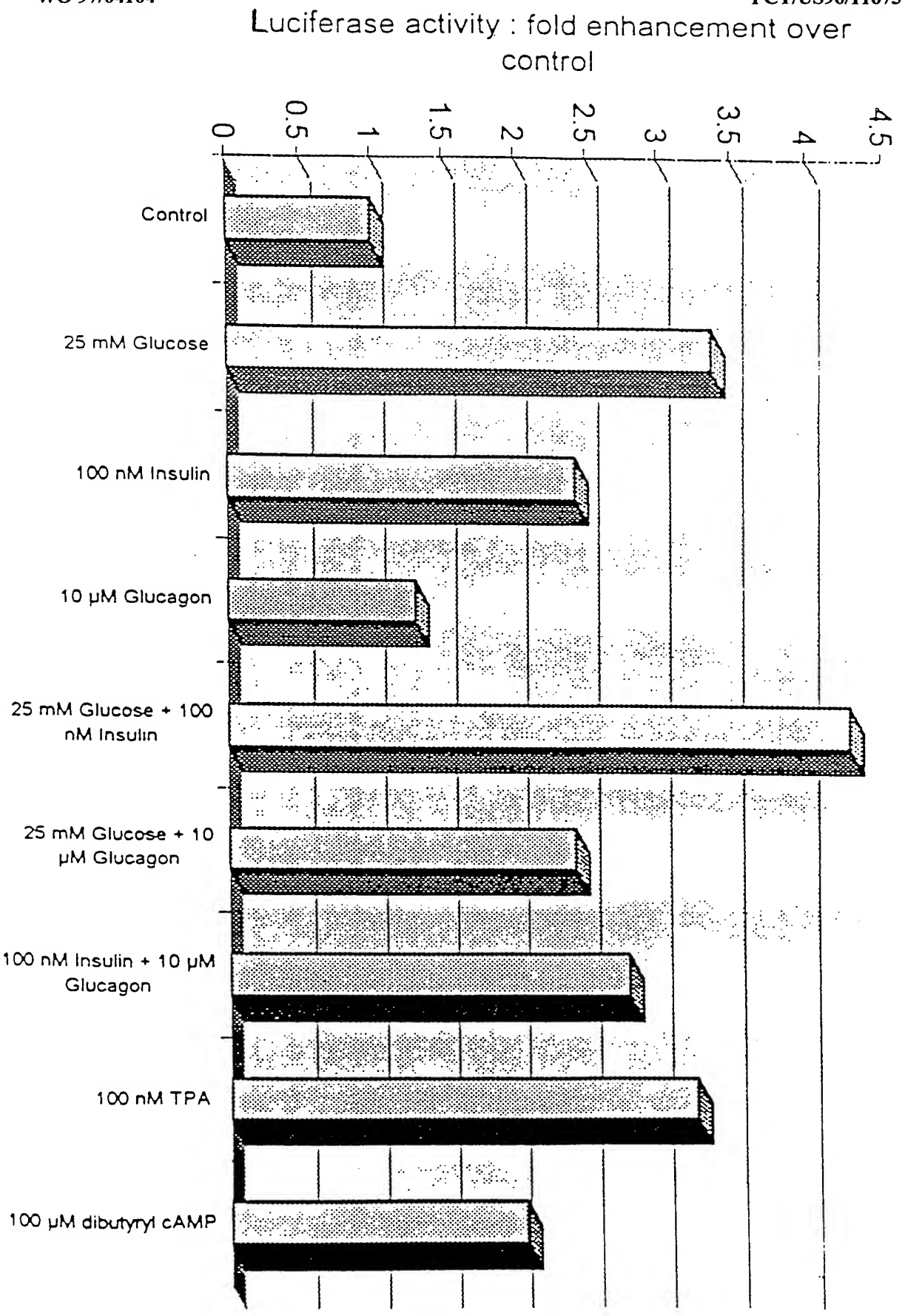

Luciferase activity : fold enhancement over control
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
Control
25 mM Glucose
100 nM Insulin
10 µM Glucagon
25 mM Glucose + 100 nM Insulin
25 mM Glucose + 10 µM Glucagon
100 nM Insulin + 10 µM Glucagon
100 nM TPA
100 µM dibutyryl cAMP

FIGURE 9

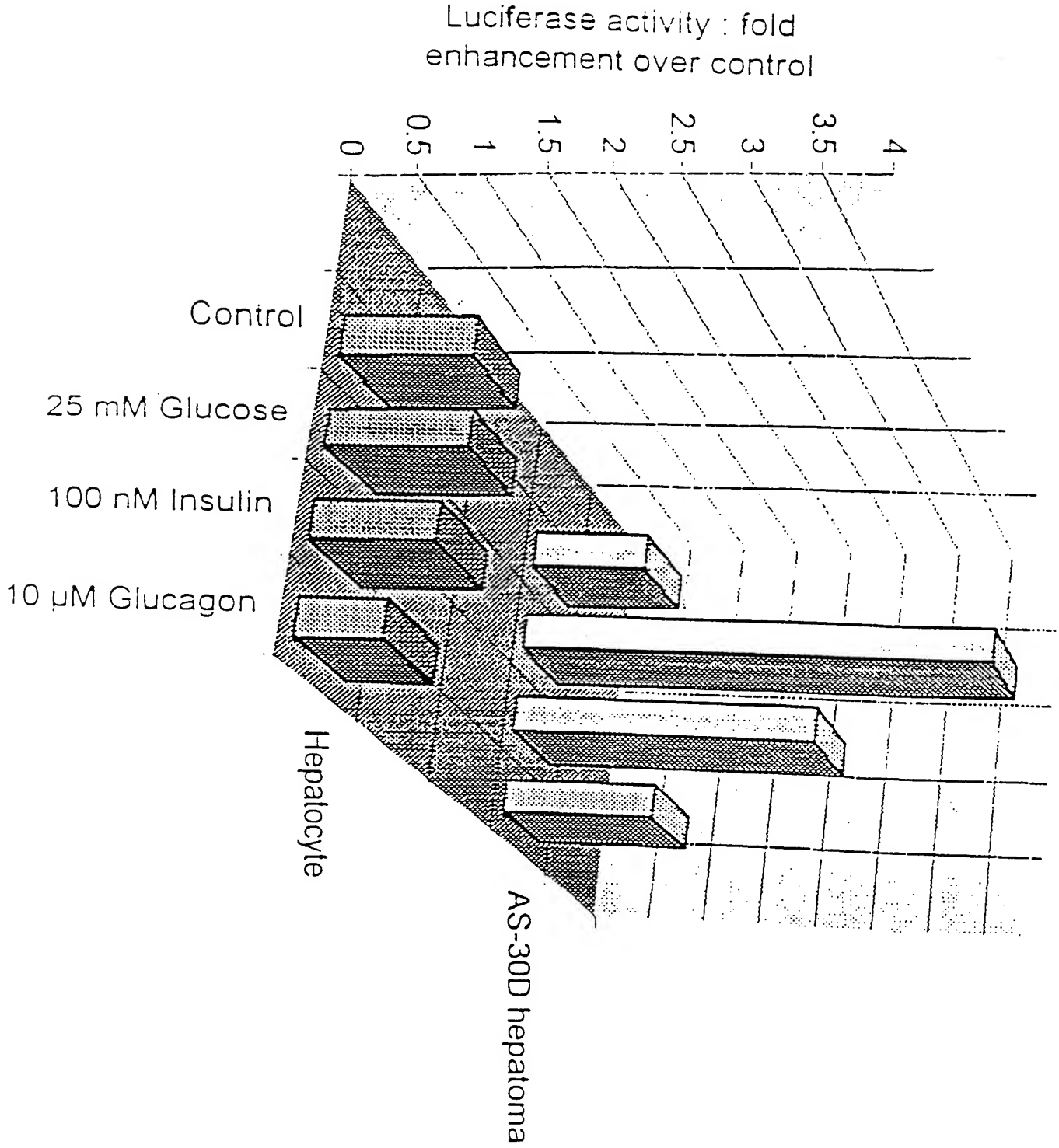

Luciferase activity : fold enhancement over control
0
0.5
1
1.5
2
2.5
3
3.5
4
Control
25 mM Glucose
100 nM Insulin
10 µM Glucagon
Hepatocyte
AS-30D hepatoma

FIGURE 10

```
HKP.SEQ: 5267 base pairs
Signal database file: MAMMAL.DAT

 (+) = Current Strand
 (-) = Opposite Strand


-4369  TCTAGAGCTCGCGGCCGCGAGCTCTAATACGACTCACTATAGGGCGTCGA
      (+)LBP-1 S00487
       (-)LBP-1 S00487
                 (-)UCE.2 S00437
                  (+)UCE.2 S00437
                                                  (-)(Spl) S01027

-4319  CTCGATCCAACTGGCCTAGAACTCACAGCCATCCTCTTGCCTCTACCTAT
              (+)multiple S01614
              (+)v-Myb S01896
              (-)c-Myb S01703
              (-)multiple S01614
                (+)LBP-1 S02121
                (+)LBP-1 S00487
                    (-)T-Ag S00972
                      (-)LBP-1 S00487
                              (+)TCF-1 S02023
                                    (+)CAP-site S00089
                                       (-)polyoma.1 S00922
                                             (-)T-Ag S00973
                                              (-)polyoma.1 S00922

-4269           GGAGTGTGGAGATTAAAGGCATGTTCTACCATGTCTTAATTTTAAAATAC
               (-)CAP-site S00089
                         (-)gamma-IRE CS S01622
                               (+)p53 S02095
                               (-)p53 S02095
                                    (+)GR S01035
                                    (+)GR S01037
                                          (-)CF1 S01946
                                                   (+)WAP US6 S01052
                                                   (-)WAP US6 S01052

-4219            CTATGGAGTGTGGAGATTAAAGGCATGTTCTACCATGTCTT
                    (-)CAP-site S00089
                              (-)gamma-IRE CS S01622
                                    (+)p53 S02095
                                    (-)p53 S02095
                                         (+)GR S01035
                                         (+)GR S01037
                                                (-)CF1 S01946

-4178  AATTTTAAA ATAGGAATATTTGTGGATTGAGGTCTTGAGCAAAATAAGA
         (+)WAP US6 S01052
         (-)WAP US6 S01052
                       (+)HNF-5 S01974
                           (+)AP-3 S01937
                            (+)multiple S01543
                              (-)CAP-site S00089
                                   (+)H-2RIIBP/T3R-al S01909
                                                    (+)GArC3 S02066
                                                    (-)GArC1 S02065

-4129    TTTTTCCCAA      GAGAGTTTCCTGAAGCCTAAGTAGACTCAGGTC
                            (-)gamma-IRE CS S01622
                                      (-)T-Ag S00972
                                                  (-)E-alpha H box S00170
                                                  (-)W-element CS S01629
                                                    (+)H4TF-2 S00742

-4086  CTTCTCATGCAGGGCCA     ATCTAGGGCCAGGAGCAGGACCAACTGGT
              (-)LVc S00040
              (-)LVc S00904
                 (-)LF-A1 S00250
                           (+)LBP-1 S00487
                               (-)LF-A1 S00250
                                        (-)LVc S00040
```

FIGURE 11A

```
                                                    (-)LVc S00904
                                                      (-)H4TF-2 S00742
                                                          (+)multiple S01614
                                                          (+)v-Myb S01896
                                                          (-)c-Myb S01703
                                                          (-)multiple S01614
                                                            (+)LBP-1 S02121
                                                            (+)LBP-1 S00487

-4040  GTGAAATCAGAAAGATGGTACTCATAGCTATTAGTCCATCTCTGGTTGAC
        (-)H4TF-1 S00741
                 (-)CAP-site S00089
                  (+)CF1 S01946
                                                  (-)CF1 S01946
                                                   (+)CAP-site S00089
                                                       (+)LBP-1 S02121
                                                       (+)LBP-1 S00487
                                                                (+)CAP-site S00089

-3990  ACTCCCAGACTCCCCTACATC TCAAGACACAGACATACGTGGCTTTTTA
         (+)SV40.11 S01003
          (-)LBP-1 S02121
          (-)LBP-1 S00487
                                      (+)TCF-1 S02023
                                                       (-)TCF-1 S02023

-3941  TGAATCCATTTTTCTGGTCTGT ATTATTTGTTCTGTGTGTTAATTTTAT
           (-)CF1 S01946
            (+)CAP-site S00089
            (+)GMCSF CS S01616
                  (+)LBP-1 S02121
                  (+)LBP-1 S00487
                               (+)HNF-5 S01974
                                   (+)GR S01035
                                   (+)GR S01037
                                       (-)TCF-1 S02023

-3892  GTCTTAAACTAAACAGAAATCCT TTAAGGAAAGAACACCCCGCCCCTCT
                    (+)TCF-1 S02023
                        (-)H4TF-1 S00741
                                     (-)IE1.2 S00794
                                         (-)GR S01035
                                         (-)GR S01037
                                             (+)CAP-site S00089
                                              (-)Sp1 S01542
                                              (-)Sp1 S00327
                                              (-)Sp1 S00064
                                               (+)(Sp1) S01187
                                               (+)AP-2 S01936
                                               (+)EARLY-SEQ1 S01081
                                               (-)JCV repeated se S01193
                                                (+)Sp1 S00801
                                                (+)Sp1 S00804
                                                (-)Sp1 S00979
                                                (-)Sp1 S00781
                                                (-)hsp70.2 S00782
                                                 (+)LSF S00975
                                                 (+)Sp1 S00799
                                                 (+)Sp1 S00802
                                                 (+)Sp1 S00803
                                                 (+)Sp1 S00805
                                                 (-)(Sp1) S01026
                                                 (-)(Sp1) S01028
                                                 (-)Sp1 S01082
                                                 (-)Sp1 S00146
                                                 (-)Sp1 S00800
                                                 (-)Sp1 S00978
                                                   (-)T-Ag S00974
                                                      (-)polyoma.1 S00922

-3843  CCTGTGTGTGTGTGTGTGTGTGTG TGTGTGTGTACACGTCTGTGTGAGT
       (-)TCF-1 S02023
        (-)CACA S00014
          (-)CACA S00014
            (-)CACA S00014
```

FIGURE 11B

```
                (-)CACA S00014
                  (-)CACA S00014
                    (-)CACA S00014
                      (-)CACA S00014
                                                    (-)TCF-1 S02023
                                                          (+)GATA-1 S00484

-3794  ATCTCGCACCCTGTAAAGGGCTTAA TAAACACGTGCTGATTGATTCCCC
            (+)CAP-site S00089
                                        (+)HNF1 S01619
                                        (+)MyoD S01905
                                        (+)multiple S01614
                                        (-)MyoD S01905
                                        (-)multiple S01614

-3745  TCCTGGAATGTGAATGTGGACTGCCA ATCTGCCAGTCTACAATGTGTGT
          (+)TEF1 S00551
           (-)CAP-site S00089
                        (-)CAP-site S00089
                            (-)TGGCA-BP S02137
                                      (-)TGGCA-BP S02137
                                        (-)LBP-1 S02121
                                        (-)LBP-1 S00487
                                         (+)CAP-site S00089

-3696  GCCTGTATGTGCTCATGGGGAGAGAGA GGGAGAAATAAAATAGACTCTA
         (+)gamma-IRE CS S01622

-3647  AGGAAGAATCTTGAGGACAGGAAAGTCA GAGCTACACACCTCACTTTTG
       (+)PEA3 S00392
                  (-)GR S00864
                       (-)IE1.2 S00794
                                              (+)CAP-site S00089
                                                  (-)IBP-1 S02119
                                                  (-)INF.1 S01152
                                                  (-)alpha-INF.2 S01153
                                                   (+)CAP-site S00089
                                                     (-)TCF-1 S02023

-3598  AGTGGGTAGCTGTCCCCTGATTTGACACA TACAGATGGGTTAGGGGATA
                                     (+)E2A S01950
                                      (+)multiple S01614
                                      (-)multiple S01614
                                       (+)CF1 S01946

-3549  TCACTGTACTCACTCCAGCCACCTCCCAGG GTTACTGGGAACTCTGTGA
         (+)gamma-IRE CS S01622
                (+)CAP-site S00089
                          (+)CAP-site S00089
                               (+)SV40.11 S01003
                                     (-)gamma-IRE CS S01622
                                         (+)LBP-1 S02121
                                         (+)LBP-1 S00487
                                          (+)H-APF-1 S01971
                                          (-)SV40.11 S01003
                                                   (-)TCF-1 S02023

-3500  GAGATCATCCCATAAAGTACCCTGTGAACAT GAGTTAGTCCTCATAAAG
           (+)CAP-site S00089
                     (+)W-element CS S01629
                           (-)TCF-1 S02023
                                                        (-)CAP-site S00089

-3451  TGGGACCAGAAAAGAGAATGGAGAATGGAGCT GAAGTGTGTGTGCAAGT
        (-)H4TF-2 S00742
           (-)LBP-1 S02121
           (-)LBP-1 S00487
               (+)TCF-1 S02023
                    (-)CAP-site S00089
                           (-)CAP-site S00089
                                      (-)CAP-site S00089
                                             (-)MBF-I S01987
                                             (-)MRE CS2 S00079

-3402  AAGTGTGTGTGAGATCCAGCTAATTGGACTCAG CTGATGGAGTGCTTGC
```

FIGURE 11C

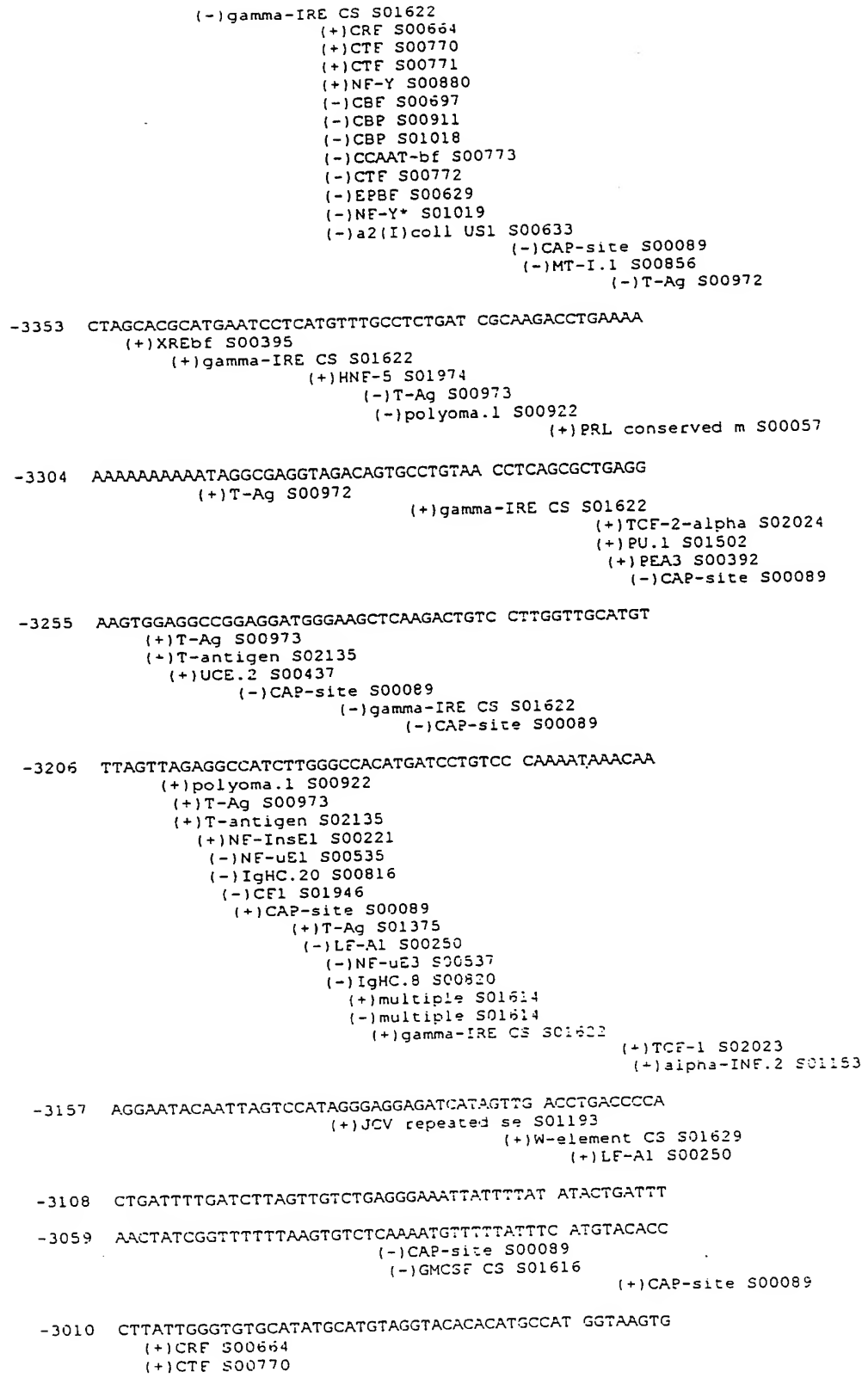

(-)gamma-IRE CS S01622
(+)CRF S00664
(+)CTF S00770
(+)CTF S00771
(+)NF-Y S00880
(-)CBF S00697
(-)CBP S00911
(-)CBP S01018
(-)CCAAT-bf S00773
(-)CTF S00772
(-)EPBF S00629
(-)NF-Y* S01019
(-)a2(I)coll US1 S00633
(-)CAP-site S00089
(-)MT-I.1 S00856
(-)T-Ag S00972
-3353 CTAGCACGCATGAATCCTCATGTTTGCCTCTGAT CGCAAGACCTGAAAA
(+)XREbf S00395
(+)gamma-IRE CS S01622
(+)HNF-5 S01974
(-)T-Ag S00973
(-)polyoma.1 S00922
(+)PRL conserved m S00057
-3304 AAAAAAAAAATAGGCGAGGTAGACAGTGCCTGTAA CCTCAGCGCTGAGG
(+)T-Ag S00972
(+)gamma-IRE CS S01622
(+)TCF-2-alpha S02024
(+)PU.1 S01502
(+)PEA3 S00392
(-)CAP-site S00089
-3255 AAGTGGAGGCCGGAGGATGGGAAGCTCAAGACTGTC CTTGGTTGCATGT
(+)T-Ag S00973
(+)T-antigen S02135
(+)UCE.2 S00437
(-)CAP-site S00089
(-)gamma-IRE CS S01622
(-)CAP-site S00089
-3206 TTAGTTAGAGGCCATCTTGGGCCACATGATCCTGTCC CAAAATAAACAA
(+)polyoma.1 S00922
(+)T-Ag S00973
(+)T-antigen S02135
(+)NF-InsE1 S00221
(-)NF-uE1 S00535
(-)IgHC.20 S00816
(-)CF1 S01946
(+)CAP-site S00089
(+)T-Ag S01375
(-)LF-A1 S00250
(-)NF-uE3 S00537
(-)IgHC.8 S00820
(+)multiple S01614
(-)multiple S01614
(+)gamma-IRE CS S01622
(+)TCF-1 S02023
(+)alpha-INF.2 S01153
-3157 AGGAATACAATTAGTCCATAGGGAGGAGATCATAGTTG ACCTGACCCCA
(+)JCV repeated se S01193
(+)W-element CS S01629
(+)LF-A1 S00250
-3108 CTGATTTTGATCTTAGTTGTCTGAGGGAAATTATTTTAT ATACTGATTT
-3059 AACTATCGGTTTTTTAAGTGTCTCAAAATGTTTTTATTTC ATGTACACC
(-)CAP-site S00089
(-)GMCSF CS S01616
(+)CAP-site S00089
-3010 CTTATTGGGTGTGCATATGCATGTAGGTACACACATGCCAT GGTAAGTG
(+)CRF S00664
(+)CTF S00770

FIGURE 11D


(+)CTF S00771
(+)NF-Y S00880
(-)CBF S00697
(-)CBP S00911
(-)CBP S01018
(-)CCAAT-bf S00773
(-)CTF S00772
(-)EPBF S00629
(-)NF-Y* S01019
(-)a2(I)coll US1 S00633
(-)MBF-I S01987
(-)MRE CS2 S00079
(+)multiple S01614
(-)multiple S01614
(+)multiple S01614
(-)multiple S01614
(-)TGGCA-BP S02137
(-)gamma-IRE CS S01622
-2961 TGTAAAGGTCAGAAGTCAATTTTCTTGAGTTGATTCTCTCCT GTGACCA
(-)AP-1 S00869
(+)CAP-site S00089
(+)gamma-IRE CS S01622
(-)CAP-site S00089
(+)multiple S01614
(-)multiple S01614
(+)CF1 S01946
-2912 CATGGGTCCTAGGGTTCACCTCAAGTAGTCGGGGTTCAGCGAC AAGCGC
(+)H4TF-2 S00742
(-)W-element CS S01629
(-)LF-A1 S00250
(+)CAP-site S00089
(-)gamma-IRE CS S01622
(-)LF-A1 S00250
-2863 CTTTACCCACCGAGCCACCTTGCCAGCACCCGAAGTGTTTCCAG AAAGG
(+)gamma-IRE CS S01622
(+)CAP-site S00089
(-)TGGCA-BP S02137
(+)W-element CS S01629
(-)CAP-site S00089
(-)gamma-IRE CS S01622
(-)H-APF-1 S01971
-2814 TCTTTTTTTTTTTCTCTTTGCTGCTTACTTTTAACCTATGCCATC AATT
(-)TCF-1 S02023
(-)TCF-1 S02023
(-)TCF-1 S02023
(-)TGGCA-BP S02137
(+)CAP-site S00089
-2765 CTGCCTCAGACTTCTGAACACCTAAAGCCTTAATCAGCCTCTGTGC CTC
(-)T-Ag S00973
(-)gamma-IRE CS S01622
(-)AP-1 S01424
(-)rPrl.A S00252
(+)CAP-site S00089
(-)T-Ag S00973
(-)polyoma.1 S00922
(-)TCF-1 S02023
(+)CAP-site S00089
-2716 ACCCTTGTCTCACTCCAGCCTTTATCTTATCTGGGAGTTCCTGTCTC TT
(+)CAP-site S00089
(+)CAP-site S00089
(+)gamma-IRE CS S01622
(-)GATA-1 S00381
(+)GATA-1 S00483
(-)GATA-1 S01590
(-)GATA-1 S00381
(+)LBP-1 S02121
(+)LBP-1 S00487
(-)SV40.11 S01003
(+)NF-kB S01498

FIGURE 11E

```
                                                    (-)gamma-IRE CS S01622
-2667  CTCCTTCAGGCCGGGTCCTTTCCTCCCATTCATGTGGAGAGCAGCTTT T
              (+)UCE.2 S00437
                 (-)W-element CS S01629
                   (+)H4TF-2 S00742
                     (-)alpha-INF.2 S01153
                       (+)IE1.2 S00794
                         (-)PU.1 S01502
                         (-)JCV repeated se S01193
                                    (+)multiple S01614
                                    (-)multiple S01614
                                              (+)CAP-site S00089
                                                      (+)GR S00864
-2618  GTCCTACAAAAGCTTTAAGCATCTCAGAGTCTGTGTCAGAAAGAGAGGA
            (+)TFIID S00435
               (+)TCF-1 S02023
                    (+)gamma-IRE CS S01622
                          (+)CAP-site S00089
                                    (-)TCF-1 S02023
                                              (+)AP-1 S00090
                                               (+)polyoma.1 S00922
-2569  GCTGGCTTATGAGGCTGTTGCAATTGGGTGAAAGACACTGGTGAACTGTG
             (+)T-Ag S00973
              (-)CAP-site S00089
                 (-)TCF-1 S02023
                  (+)C/EBP S00121
                  (+)NF-IL6 S02000
                       (+)multiple S01614
                       (-)multiple S01614
                         (+)CRF S00664
                         (+)CTF S00770
                         (+)CTF S00771
                         (+)NF-Y S00880
                         (-)CBF S00697
                         (-)CBP S00911
                         (-)CBP S01018
                         (-)CCAAT-bf S00773
                         (-)CTF S00772
                         (-)EPBF S00629
                         (-)NF-Y* S01019
                         (-)a2(I)coll US1 S00633
                                    (+)LBP-1 S02121
                                    (+)LBP-1 S00487
                                          (-)CAP-site S00089
                                             (-)TCF-1 S02023
-2519    AGGCAGACCAATGGGAAGGGTTTGAGAACTAATATAGAAAATGAAAGTCT
       (+)T-Ag S00973
              (+)CBF S00697
              (+)CBP S00911
              (+)CBP S01018
              (+)CCAAT-bf S00773
              (+)CTF S00772
              (+)EPBF S00629
              (+)NF-Y* S01019
              (+)a2(I)coll US1 S00633
              (-)CRF S00664
              (-)CTF S00770
              (-)CTF S00771
              (-)NF-Y S00880

                                         (+)IFN-inducible C S01550
                                           (-)CAP-site S00089
                                            (+)alpha-INF.2 S01153
                                            (-)GMCSF CS S01616
                                             (+)Ad-conserved-se S01090
-2471    CTCCTTTGTGTCGTATAATCATATGTGACATCACTAAATCATCTACTA
         (-)alpha-INF.2 S01153
           (-)TCF-1 S02023
                    (+)GATA-1 S00477
                         (+)multiple S01614
                         (-)multiple S01614
```

FIGURE 11F

```
                                    (-)TCR-beta decame S00185

-2421  AC TTACACAATAAATACCTACATGGTGCCTACCATGTGATAGAGCGCCC
        (-)C/EBP S00121
        (-)NF-IL6 S02000
              (+)GH-CSE2 S00192
                    (+)CF1 S01946
                          (-)T-Ag S00972
                               (-)CF1 S01946
                                (+)multiple S01614
                                (-)multiple S01614
                                    (+)GATA-1 S00381
                                    (-)GATA-1 S00481
                                          (+)GCF S01964
                                            (-)T-Ag S00974

-2372  CCA CACGAGGTACTGCAGATAGAAGGGAATGATATAGACGCAGATGCTT
          (-)uteroglobin HS- S01190
              (-)gamma-IRE CS S01622
                    (+)GATA-1 S00381
                         (+)alpha-INF.2 S01153
                            (-)CAP-site S00089
                              (+)Ad-conserved-se S01090
                                        (+)E2A S01950
                                         (+)multiple S01614
                                         (-)multiple S01614

-2323  ATTC AGACAGGTAGGACAGAATGGATATAACACTTAGAAAAGGACCCGG
                   (-)GR S00864
                        (-)CAP-site S00089
                           (-)Pit-1 S00018
                                       (+)TCF-1 S02023
                                        (+)alpha-INF.2 S01153
                                          (+)W-element CS S01629
                                           (-)H4TF-2 S00742

-2274  GTGTG GTGGCACGGTGGCACATACCTTAGATCCCACCACCCGGGTGTTG
             (+)TGGCA-BP S02137
                    (+)TGGCA-BP S02137
                                  (+)HC3 S00243
                                                    (+)T-Ag S00974

-2225  GGGCTG AGGCAGATGACTTTTGTTTGTTTTTTTGTTTTCAGTTTTGCTT
       (-)CAP-site S00089
             (+)T-Ag S00973
                (+)E2A S01950
                 (+)multiple S01614
                 (-)multiple S01614
                        (-)TCF-1 S02023
                            (+)HNF-5 S01974
                                          (+)CAP-site S00089
                                                  (-)TCF-1 S02023

-2176  TTTTTCA AGACAGGGTCTCTCCGTGCAGCACTGCCTGTCCTGGAACTCG
                                       (+)GR S00864
                                           (+)gamma-IRE CS S01622

-2127  CTTTGCTG GCCACGTTTGTGGCCTTGAACTCACAAAGTGCTGGTGCCTG
       (-)TCF-1 S02023
                           (+)gamma-IRE CS S01622
                                    (+)TCF-1 S02023
                                     (-)CAP-site S00089

-2078  GCTGTAAAA TTAATTTCTCTCTCCCTCTCTCCCCCCCTCCCCCACCTCT
                            (-)polyoma.1 S00922
                                 (-)JCV repeated se S01193
                                  (-)JCV repeated se S01193
                                    (+)AP-2 S00346
                                     (+)AP-2 S01936
                                     (-)JCV repeated se S01193
                                      (-)H4TF1 S01969
                                              (+)CAP-site S00089
                                                (-)polyoma.1 S00922

-2029  CTCGCTACTT GCTTGGTAGACCAGACTGGCTTCGAACTCAGAGATTTGC
```

FIGURE 11G

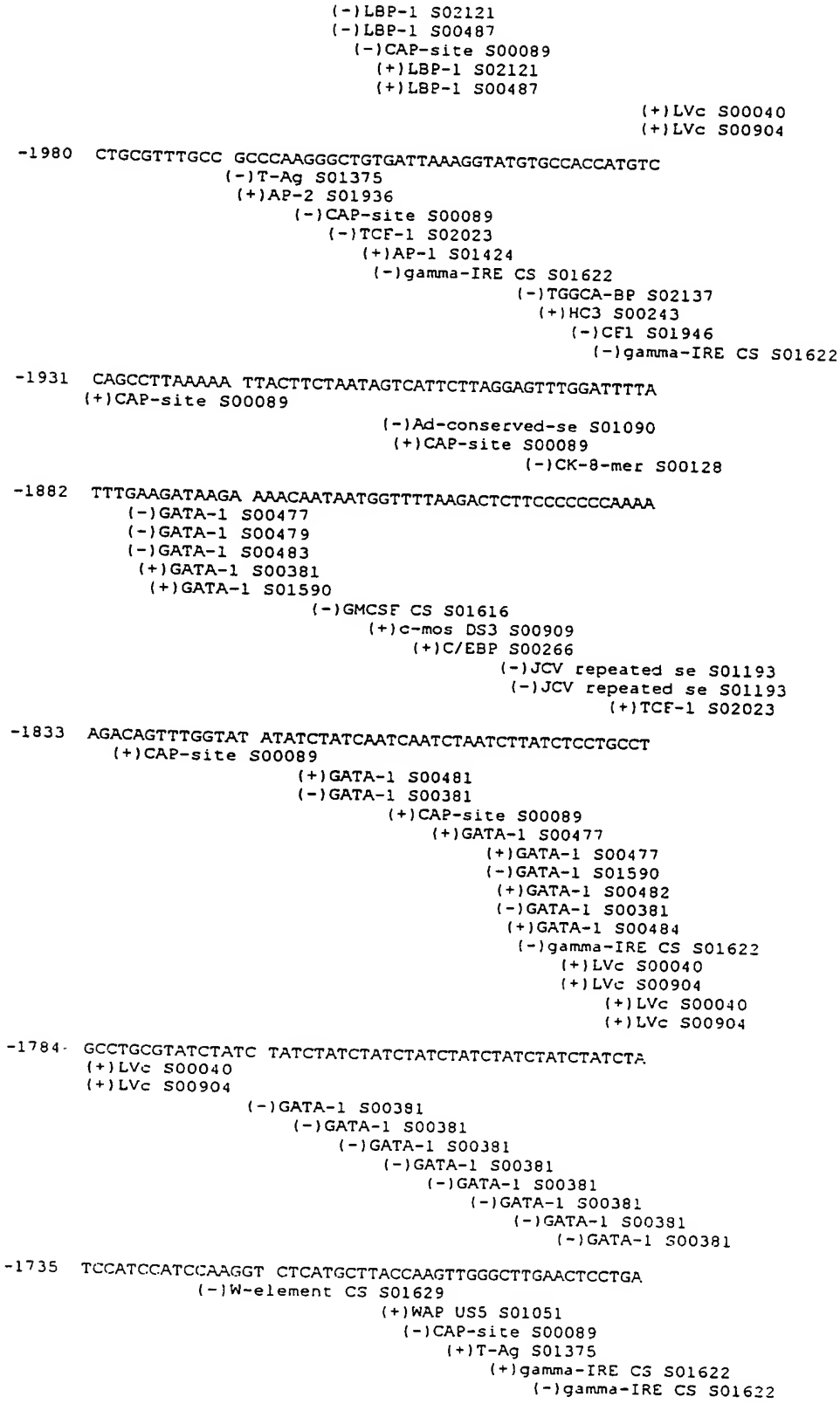

(-)LBP-1 S02121
(-)LBP-1 S00487
(-)CAP-site S00089
(+)LBP-1 S02121
(+)LBP-1 S00487
(+)LVc S00040
(+)LVc S00904
-1980 CTGCGTTTGCC GCCCAAGGGCTGTGATTAAAGGTATGTGCCACCATGTC
(-)T-Ag S01375
(+)AP-2 S01936
(-)CAP-site S00089
(-)TCF-1 S02023
(+)AP-1 S01424
(-)gamma-IRE CS S01622
(-)TGGCA-BP S02137
(+)HC3 S00243
(-)CF1 S01946
(-)gamma-IRE CS S01622
-1931 CAGCCTTAAAAA TTACTTCTAATAGTCATTCTTAGGAGTTTGGATTTTA
(+)CAP-site S00089
(-)Ad-conserved-se S01090
(+)CAP-site S00089
(-)CK-8-mer S00128
-1882 TTTGAAGATAAGA AAACAATAATGGTTTTAAGACTCTTCCCCCCCAAAA
(-)GATA-1 S00477
(-)GATA-1 S00479
(-)GATA-1 S00483
(+)GATA-1 S00381
(+)GATA-1 S01590
(-)GMCSF CS S01616
(+)c-mos DS3 S00909
(+)C/EBP S00266
(-)JCV repeated se S01193
(-)JCV repeated se S01193
(+)TCF-1 S02023
-1833 AGACAGTTTGGTAT ATATCTATCAATCAATCTAATCTTATCTCCTGCCT
(+)CAP-site S00089
(+)GATA-1 S00481
(-)GATA-1 S00381
(+)CAP-site S00089
(+)GATA-1 S00477
(+)GATA-1 S00477
(-)GATA-1 S01590
(+)GATA-1 S00482
(-)GATA-1 S00381
(+)GATA-1 S00484
(-)gamma-IRE CS S01622
(+)LVc S00040
(+)LVc S00904
(+)LVc S00040
(+)LVc S00904
-1784 GCCTGCGTATCTATC TATCTATCTATCTATCTATCTATCTATCTATCTA
(+)LVc S00040
(+)LVc S00904
(-)GATA-1 S00381
(-)GATA-1 S00381
(-)GATA-1 S00381
(-)GATA-1 S00381
(-)GATA-1 S00381
(-)GATA-1 S00381
(-)GATA-1 S00381
(-)GATA-1 S00381
-1735 TCCATCCATCCAAGGT CTCATGCTTACCAAGTTGGGCTTGAACTCCTGA
(-)W-element CS S01629
(+)WAP US5 S01051
(-)CAP-site S00089
(+)T-Ag S01375
(+)gamma-IRE CS S01622
(-)gamma-IRE CS S01622

FIGURE 11H

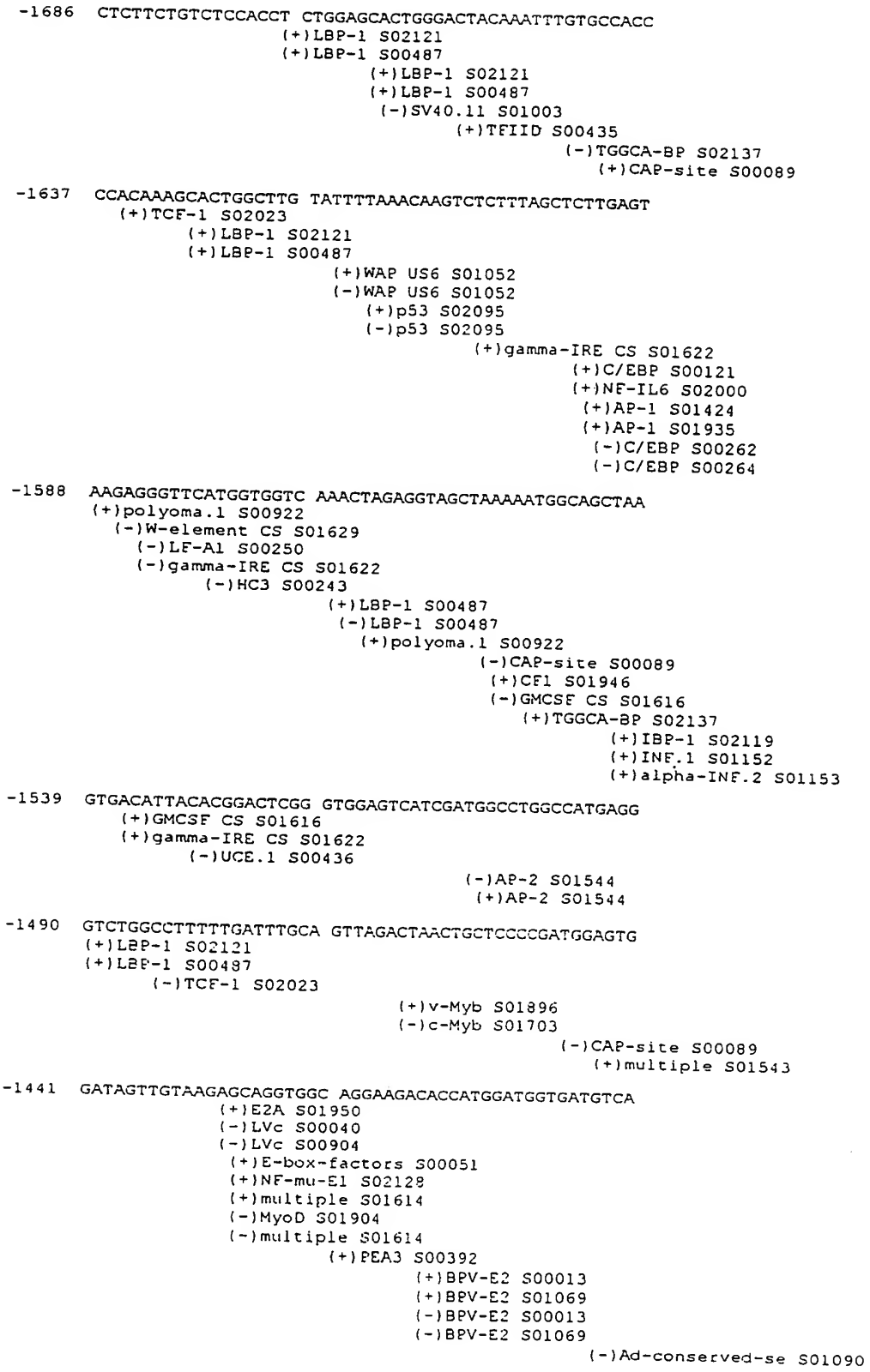

-1686 CTCTTCTGTCTCCACCT CTGGAGCACTGGGACTACAAATTTGTGCCACC
(+)LBP-1 S02121
(+)LBP-1 S00487
(+)LBP-1 S02121
(+)LBP-1 S00487
(-)SV40.11 S01003
(+)TFIID S00435
(-)TGGCA-BP S02137
(+)CAP-site S00089
-1637 CCACAAAGCACTGGCTTG TATTTTAAACAAGTCTCTTTAGCTCTTGAGT
(+)TCF-1 S02023
(+)LBP-1 S02121
(+)LBP-1 S00487
(+)WAP US6 S01052
(-)WAP US6 S01052
(+)p53 S02095
(-)p53 S02095
(+)gamma-IRE CS S01622
(+)C/EBP S00121
(+)NF-IL6 S02000
(+)AP-1 S01424
(+)AP-1 S01935
(-)C/EBP S00262
(-)C/EBP S00264
-1588 AAGAGGGTTCATGGTGGTC AAACTAGAGGTAGCTAAAAATGGCAGCTAA
(+)polyoma.1 S00922
(-)W-element CS S01629
(-)LF-A1 S00250
(-)gamma-IRE CS S01622
(-)HC3 S00243
(+)LBP-1 S00487
(-)LBP-1 S00487
(+)polyoma.1 S00922
(-)CAP-site S00089
(+)CF1 S01946
(-)GMCSF CS S01616
(+)TGGCA-BP S02137
(+)IBP-1 S02119
(+)INF.1 S01152
(+)alpha-INF.2 S01153
-1539 GTGACATTACACGGACTCGG GTGGAGTCATCGATGGCCTGGCCATGAGG
(+)GMCSF CS S01616
(+)gamma-IRE CS S01622
(-)UCE.1 S00436
(-)AP-2 S01544
(+)AP-2 S01544
-1490 GTCTGGCCTTTTTGATTTGCA GTTAGACTAACTGCTCCCCGATGGAGTG
(+)LBP-1 S02121
(+)LBP-1 S00487
(-)TCF-1 S02023
(+)v-Myb S01896
(-)c-Myb S01703
(-)CAP-site S00089
(+)multiple S01543
-1441 GATAGTTGTAAGAGCAGGTGGC AGGAAGACACCATGGATGGTGATGTCA
(+)E2A S01950
(-)LVc S00040
(-)LVc S00904
(+)E-box-factors S00051
(+)NF-mu-E1 S02128
(+)multiple S01614
(-)MyoD S01904
(-)multiple S01614
(+)PEA3 S00392
(+)BPV-E2 S00013
(+)BPV-E2 S01069
(-)BPV-E2 S00013
(-)BPV-E2 S01069
(-)Ad-conserved-se S01090

FIGURE 11I

```
                                                    (-)alpha-INF.2 S01153
                                                     (+)multiple S01614
                                                     (+)GMCSF CS S01616
                                                     (-)multiple S01614

-1392  TTTGTGGAGACAACTGGTAAAGG AAAAAAAAAACCTAAGAAGTTCAGCT
                (+)multiple S01614
                (+)v-Myb S01896
                (-)c-Myb S01703
                (-)multiple S01614
                  (+)LBP-1 S02121
                  (+)LBP-1 S00487
                                                        (+)CAP-site S00089

-1343  TAGTACATGTTAAGTGTGAGGTGC TAGTCACCTCTGCTGGGATGACAAA
                        (-)CAP-site S00089
                              (-)AP-1 S02100
                              (-)LBP-1 S00487
                                   (+)CAP-site S00089
                                     (-)polyoma.1 S00922
                                           (-)SV40.11 S01003
                                             (-)CAP-site S00089
                                                   (-)c-mos DS1 S00907

-1294  CCATAGCTGGCTAAGAAAGTTAGAA GCCCTGGGGGAAAGTTCTTGCTTT
                                  (-)AP-2 S00346
                                  (-)AP-2 S01936
                                   (-)AP-2 S00346
                                   (-)AP-2 S01936
                                     (-)SV40.11 S01003
                                          (-)IE1.2 S00794
                                                 (-)TCF-1 S02023
                                                  (-)TFIID S00435

-1245  GTAGATTAGGTTTGTGAGTACCTGTG TGCAGATGGTGTAACTGCTGTCG
                        (+)W-element CS S01629
                             (-)TCF-1 S02023
                                    (+)E2A S01950
                                     (+)multiple S01614
                                     (-)multiple S01614
                                      (+)CF1 S01946
                                            (+)W-element CS S01629
                                              (+)v-Myb S01896
                                              (-)c-Myb S01703

-1196  ACAGTGCTGGGAGGAATCGCCCATCGA GGGAATAGATGAACCACACCAG
                (-)SV40.11 S01003
                  (+)JCV repeated se S01193
                    (+)PU.1 S01502
                            (-)T-Ag S01375
                                    (-)vaccinia-term-s S01197
                                           (+)LF-A1 S00250
                                                  (-)LBP-1 S02121
                                                  (-)LBP-1 S00487

-1147  AGAACATGGTAGAAGCGGCCCAGCAGAG CAACGTGGGCTGGGGTGTACT
       (-)GR S01035
       (-)GR S01037
          (+)CF1 S01946
                      (+)GCF S01964
                       (-)UCE.2 S00437
                         (-)T-Ag S01375
                          (+)SV40.11 S01003
                                        (+)T-Ag S01375
                                         (-)AP-2 S01936
                                         (-)CAP-site S00089
                                          (-)AP-2 S01936
                                            (-)SV40.11 S01003
                                              (-)CAP-site S00089

-1098  TCAGTCGGCAGAGTGCTTTGATCTCCAGT AGTGGCCCATGCCACTCCAG
                (-)CAP-site S00089
                 (-)MT-I.1 S00856
                      (-)TCF-1 S02023
                           (-)gamma-IRE CS S01622
```

FIGURE 11J

```
                                    (-)AP-1 S01465
                                    (-)LBP-1 S02121
                                    (-)LBP-1 S00487
                                            (+)LF-A1 S00250
                                              (-)T-Ag S01375
                                               (+)AP-2 S01936
                                                   (-)TGGCA-BP S02137
                                                      (+)CAP-site S00089
                                                          (-)LBP-1 S02121
                                                          (-)LBP-1 S00487
                                                            (+)polyoma.1 S00922
                                                             (-)ApoE B1 S00258
                                                             (-)CAP-site S00089

-1049  AGGTGGGGAGAGAGCTTGGGGAGCGAGACT GTTTGGAAATGGTAGGCCC
                 (-)TFII-I S01892
                                                (+)NF-IL6 S02000
                                                    (-)CAP-site S00089
                                                     (+)CF1 S01946
                                                     (-)GMCSF CS S01616
                                                           (+)T-Ag S00972

-1000  TGTCTTCTTTCCATGTAACTTTCCAACTCCC AGGTTTCCCATTCTCCAC
                (+)IE1.2 S00794
                (-)SV40.13 S01006
                (-)SV40.16 S00989
                (-)SV40.6 S00988
                    (-)CF1 S01946
                       (+)W-element CS S01629
                            (+)IE1.2 S00794
                            (-)SV40.13 S01006
                            (-)SV40.16 S00989
                            (-)SV40.6 S00988
                                 (+)CAP-site S00089
                                                    (+)CAP-site S00089
                                                          (+)HC3 S00243

-951   CAGCAACAACCTCATGCCATTTGAGGTGCTAC TTCAATATCGCTGGCGT
                (+)CAP-site S00089
                  (-)gamma-IRE CS S01622
                        (-)TGGCA-BP S02137
                          (-)CF1 S01946
                           (+)multiple S01614
                           (+)GMCSF CS S01616
                           (-)multiple S01614
                                (-)CAP-site S00089

-902   CTACTCATCTATGTGAACTTAAGAGTCTACTCA TCTATGTGACTAAGAG
       (-)AP-1 S02100
                                  (-)AP-1 S02100
                                       (+)CAP-site S00089
                                              (+)AP-1 S00143
                                              (+)AP-1 S01452
                                               (+)AP-1 S01424
                                               (+)AP-1 S01935
                                               (+)AP-1 S00982
                                               (+)PEA1 S00390
                                               (-)PEA1 S01595
                                                        (+)uteroglobin HS- S01190

-853   TCTGGTGTCAGGCGTGAGCTGAGGTAGAGGTGGG CTCTTCTCAGCCTCT
       (+)LBP-1 S02121
       (+)LBP-1 S00487
              (-)glucagon-G3A S02046
                    (-)CAP-site S00089
                          (+)polyoma.1 S00922
                           (-)CAP-site S00089
                                                (+)CAP-site S00089
                                                  (-)T-Ag S00973 -
                                                   (-)polyoma.1 S00922
                                                       (+)(TFIID/TBF) S00783
                                                       (+)TFIID S00087
                                                       (+)Ad2MLP US.3 S00613

-804   ATAAACCAATTCACACCACTTGAGCCAAGCAGTTA CACATGCACTTTCT
```

FIGURE 11K

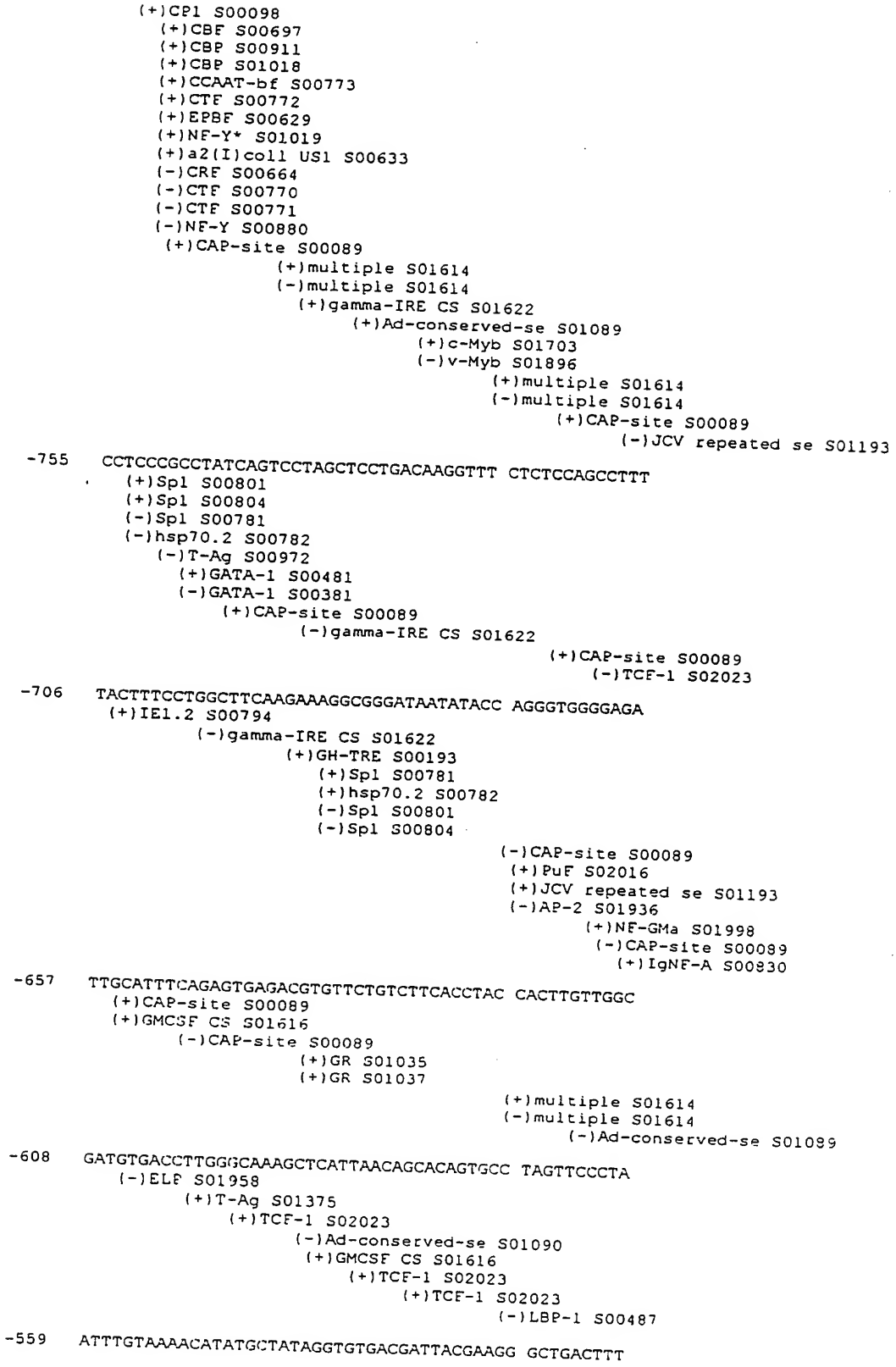

(+)CP1 S00098
(+)CBF S00697
(+)CBP S00911
(+)CBP S01018
(+)CCAAT-bf S00773
(+)CTF S00772
(+)EPBF S00629
(+)NF-Y* S01019
(+)a2(I)coll US1 S00633
(-)CRF S00664
(-)CTF S00770
(-)CTF S00771
(-)NF-Y S00880
(+)CAP-site S00089
(+)multiple S01614
(-)multiple S01614
(+)gamma-IRE CS S01622
(+)Ad-conserved-se S01089
(+)c-Myb S01703
(-)v-Myb S01896
(+)multiple S01614
(-)multiple S01614
(+)CAP-site S00089
(-)JCV repeated se S01193
-755 CCTCCCGCCTATCAGTCCTAGCTCCTGACAAGGTTT CTCTCCAGCCTTT
(+)Sp1 S00801
(+)Sp1 S00804
(-)Sp1 S00781
(-)hsp70.2 S00782
(-)T-Ag S00972
(+)GATA-1 S00481
(-)GATA-1 S00381
(+)CAP-site S00089
(-)gamma-IRE CS S01622
(+)CAP-site S00089
(-)TCF-1 S02023
-706 TACTTTCCTGGCTTCAAGAAAGGCGGGATAATATACC AGGGTGGGGAGA
(+)IE1.2 S00794
(-)gamma-IRE CS S01622
(+)GH-TRE S00193
(+)Sp1 S00781
(+)hsp70.2 S00782
(-)Sp1 S00801
(-)Sp1 S00804
(-)CAP-site S00089
(+)PuF S02016
(+)JCV repeated se S01193
(-)AP-2 S01936
(+)NF-GMa S01998
(-)CAP-site S00089
(+)IgNF-A S00830
-657 TTGCATTTCAGAGTGAGACGTGTTCTGTCTTCACCTAC CACTTGTTGGC
(+)CAP-site S00089
(+)GMCSF CS S01616
(-)CAP-site S00089
(+)GR S01035
(+)GR S01037
(+)multiple S01614
(-)multiple S01614
(-)Ad-conserved-se S01089
-608 GATGTGACCTTGGGCAAAGCTCATTAACAGCACAGTGCC TAGTTCCCTA
(-)ELF S01958
(+)T-Ag S01375
(+)TCF-1 S02023
(-)Ad-conserved-se S01090
(+)GMCSF CS S01616
(+)TCF-1 S02023
(+)TCF-1 S02023
(-)LBP-1 S00487
-559 ATTTGTAAAACATATGCTATAGGTGTGACGATTACGAAGG GCTGACTTT

FIGURE 11L

```
(-)TFIID S00435
        (+)multiple S01614
        (-)multiple S01614
                    (+)EivF/CREB S00104
                    (-)E4F1 S01258
                    (-)E4F1 S01250
                    (-)E4F1 S01251
                     (-)ATF S02101
                     (-)CREB S00489
                                          (-)TCF-1 S02023
                                           (+)C/EBP S00121
                                           (+)NF-IL6 S02000
                                            (-)TFIID S00435

-510  TGTAATGGCTTTGCTTCAGGGATCTGCAGACTCGTTGAGCC ACAATTAG
       (-)GMCSF CS S01616
            (-)TCF-1 S02023
                                      (-)HNF1 S01619
                                          (-)CAP-site S00089

-461  GATGAGAATCAAGGTGCTTCAGACTTGTGACAGGGCACTGGC GGCCCCT
          (-)gamma-IRE CS S01622
               (-)CAP-site S00089
               (-)W-element CS S01629
                                  (+)LBP-1 S02121
                                  (+)LBP-1 S00487
                                        (-)UCE.2 S00437
                                         (-)T-Ag S00974

-412  CACATGATCCTCAGATACCAGATTGTGGCGTGTGCTGCTAGGA TCACTT
      (+)multiple S01614
      (-)multiple S01614
        (+)gamma-IRE CS S01622
                (+)NF-GMb S01999
                      (-)LBP-1 S02121
                      (-)LBP-1 S00487
                        (-)CAP-site S00089
                                          (-)IBP-1 S02119
                                          (-)INF.1 S01152
                                          (-)alpha-INF.2 S01153
                                           (+)multiple S01614
                                           (-)multiple S01614

-363  GTCTTTCCAGTCTCCCAACATCTCTTGGGTCCGTGATCACGCGC CCCCC
        (+)IE1.2 S00794
        (-)SV40.13 S01006
        (-)SV40.16 S00989
        (-)SV40.6 S00988
         (-)H-APF-1 S01971
            (-)LBP-1 S02121
            (-)LBP-1 S00487
             (+)CAP-site S00089
                       (+)CAP-site S00089
                                (+)H4TF-2 S00742
                                 (+)UCE.1 S00436
                                          (+)AP-2 S00346
                                           (+)AP-2 S00346
                                            (+)AP-2 S01936
                                            (-)JCV repeated se S01193
                                             (+)AP-2 S01936
                                             (-)PuF S02016

-314  ACCCGAAGCCCAGCCTGACGCGGCGGTGGCTCATGCGCCCTGGAG TCCC
             (-)T-Ag S01375
              (+)SV40.11 S01003
                (+)CAP-site S00089
                     (-)CREB S00489
                        (-)TTR inverted re S01112
                         (-)GCF S01964
                                (-)gamma-IRE CS S01622
                                     (+)MRE CS2 S00079
                                          (+)gamma-IRE CS S01622

-265  GGGCTCTAGCCACGGAACACACGTCCCAACTCTGGCGCCCGGCTCC GCC
         (+)LBP-1 S00487
```

FIGURE 11M

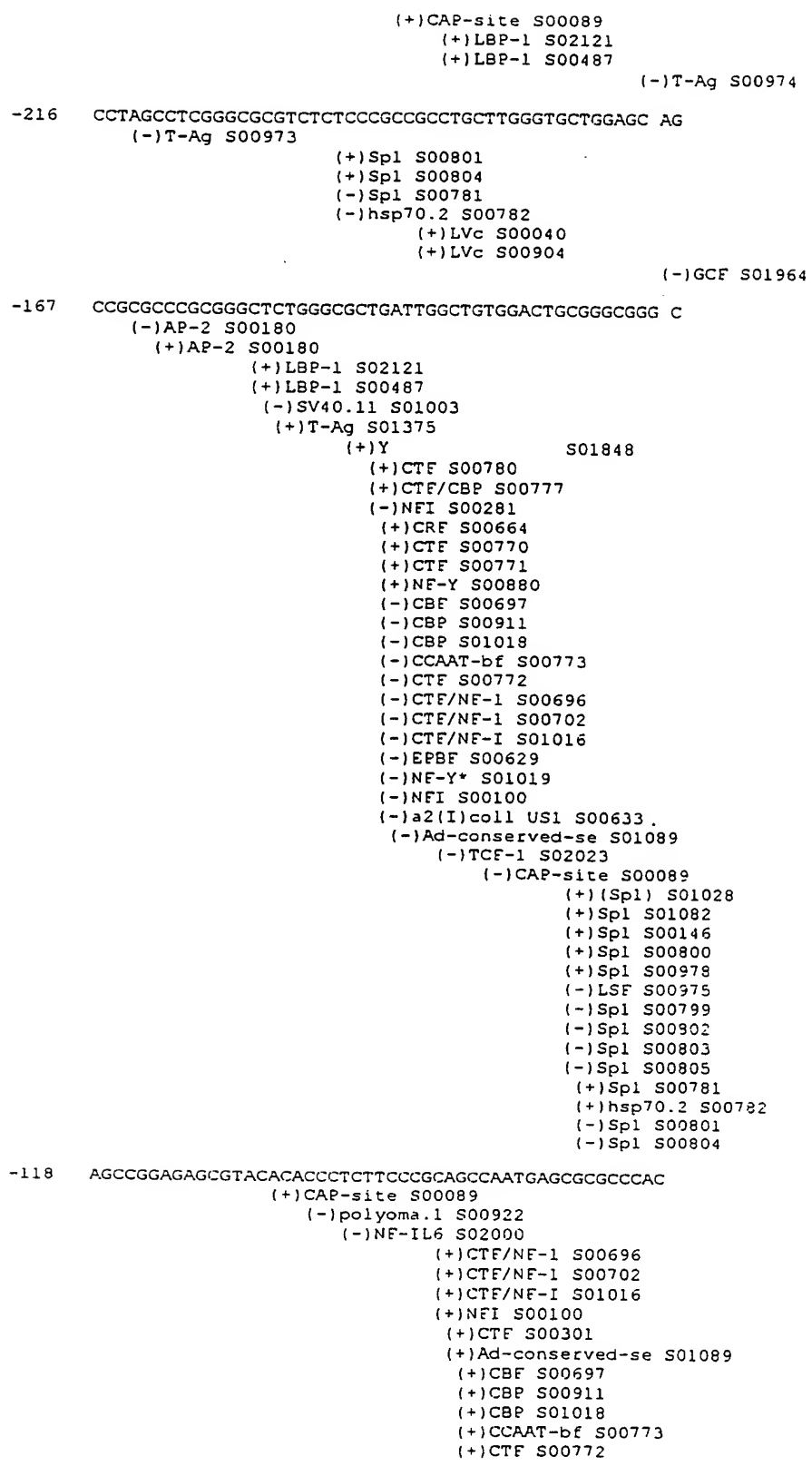

(+)CAP-site S00089
(+)LBP-1 S02121
(+)LBP-1 S00487
(-)T-Ag S00974
-216 CCTAGCCTCGGGCGCGTCTCTCCCGCCGCCTGCTTGGGTGCTGGAGC AG
(-)T-Ag S00973
(+)Sp1 S00801
(+)Sp1 S00804
(-)Sp1 S00781
(-)hsp70.2 S00782
(+)LVc S00040
(+)LVc S00904
(-)GCF S01964
-167 CCGCGCCCGCGGGCTCTGGGCGCTGATTGGCTGTGGACTGCGGGCGGG C
(-)AP-2 S00180
(+)AP-2 S00180
(+)LBP-1 S02121
(+)LBP-1 S00487
(-)SV40.11 S01003
(+)T-Ag S01375
(+)Y S01848
(+)CTF S00780
(+)CTF/CBP S00777
(-)NFI S00281
(+)CRF S00664
(+)CTF S00770
(+)CTF S00771
(+)NF-Y S00880
(-)CBF S00697
(-)CBP S00911
(-)CBP S01018
(-)CCAAT-bf S00773
(-)CTF S00772
(-)CTF/NF-1 S00696
(-)CTF/NF-1 S00702
(-)CTF/NF-I S01016
(-)EPBF S00629
(-)NF-Y* S01019
(-)NFI S00100
(-)a2(I)coll US1 S00633
(-)Ad-conserved-se S01089
(-)TCF-1 S02023
(-)CAP-site S00089
(+)(Sp1) S01028
(+)Sp1 S01082
(+)Sp1 S00146
(+)Sp1 S00800
(+)Sp1 S00978
(-)LSF S00975
(-)Sp1 S00799
(-)Sp1 S00902
(-)Sp1 S00803
(-)Sp1 S00805
(+)Sp1 S00781
(+)hsp70.2 S00782
(-)Sp1 S00801
(-)Sp1 S00804
-118 AGCCGGAGAGCGTACACACCCTCTTCCCGCAGCCAATGAGCGCGCCCAC
(+)CAP-site S00089
(-)polyoma.1 S00922
(-)NF-IL6 S02000
(+)CTF/NF-1 S00696
(+)CTF/NF-1 S00702
(+)CTF/NF-I S01016
(+)NFI S00100
(+)CTF S00301
(+)Ad-conserved-se S01089
(+)CBF S00697
(+)CBP S00911
(+)CBP S01018
(+)CCAAT-bf S00773
(+)CTF S00772

FIGURE 11N

```
                                        (+)EPBF S00629
                                        (+)NF-Y* S01019
                                        (+)a2(I)coll US1 S00633
                                        (-)CRF S00664
                                        (-)CTF S00770
                                        (-)CTF S00771
                                        (-)NF-Y S00880
                                          (+)Ad-conserved-se S01090
                                               (+)GCF S01964
                                                   (-)T-Ag S01375
                                                        (+)CREB S00489

-69   GTCACTGTCTTGGGCGGCCCAAAGAGCCGGCAGCCCCTCAATAAGCCACA
                 (+)T-Ag S01375
                  (+)(Sp1) S01028
                  (+)Sp1 S01082
                  (+)Sp1 S00146
                  (+)Sp1 S00800
                  (+)Sp1 S00978
                  (-)LSF S00975
                  (-)Sp1 S00799
                  (-)Sp1 S00802
                  (-)Sp1 S00803
                  (-)Sp1 S00805
                    (+)GCF S01964
                     (-)UCE.2 S00437
                       (-)T-Ag S01375
                           (+)TCF-1 S02023
                                    (+)AP-2 S01544
                                      (+)CAP-site S00089
                                        (-)T-Ag S00974
                                                        (+)multiple S01614
                                                        (-)multiple S01614

-19    TTGTTGCACCAACTCCAGTGCTAGAGTCTCAGGACACCACAGGCTACAC
               (+)CAP-site S00089
                (-)gamma-IRE CS S01622
                    (-)LBP-1 S02121
                    (-)LBP-1 S00487
                          (-)LBP-1 S00487
                                   (+)gamma-IRE CS S01622
                                    (-)GR S00864
                                          (+)TCF-1 S02023

+31    G GAGTTATCCCGCTTAGGAGACCCGAAGGCAGGAGCATCACTCCAGTGA
                                  (-)LVc S00040
                                  (-)LVc S00904
                                           (+)CAP-site S00089
                                               (-)LBP-1 S02121
                                               (-)LBP-1 S00487

+80    CT CTGATAAGGTGCGATCGCCCGAGAGGAACAGAACTGTCATTTTTGCG
          (+)GATA-1 S00381
           (+)GATA-1 S01590
              (-)CAP-site S00089
                            (+)AP-1 S00090
                             (+)polyoma.1 S00922
                              (+)PU.1 S01502
                                 (+)LVa S00038
                                 (+)LVa S00900
                                  (+)TCF-1 S02023
                                      (-)CAP-site S00089
                                            (-)Ad-conserved-se S01090
                                            (-)alpha-INF.2 S01153
                                             (+)CAP-site S00089
                                             (+)GMCSF CS S01616

+129   AAG TTGAGCCTTACGGATCCCGTGGGCGAAGTTAGCGACGGGACGCTGA
                          (-)AP-2 S01936
                           (-)AP-2 S00346
                              (+)T-Ag S01375
                                          (-)SIF S02021

+178   GCAA CTAGACCGGACGGCAGGAGTGAGACTTAGGTGCCTTCTAGTAGTT
          (+)LBP-1 S00487
```

FIGURE 11O

```
          (-)LBP-1 S00487
                    (-)LVc S00040
                    (-)LVc S00904
                       (-)CAP-site S00089
                           (+)AP-1 S01458
                                                (+)LBP-1 S00487
                                                 (-)LBP-1 S00487

+227   GTGAC TTAAAAAAAAAAAAAAAAAGGAAAAGAAAAAAGGAGGAAAACCTG
                          (+)TCF-1 S02023
                           (+)alpha-INF.2 S01153
                                (+)TCF-1 S02023
                                       (+)TCF-1 S02023
                                        (+)alpha-INF.2 S01153
                                            (+)PU.1 S01502
                                                    (-)TCF-1 S02023

+276   TTTCTG GAAACGCGAGGCCCTCAGCTGGTGAGCCATCGTGGTTAAGCTT
                     (+)AP-2 S01544
                       (+)T-Ag S00973
                       (+)T-antigen S02135
                               (+)multiple S01614
                               (-)GT-2B S02113
                               (-)multiple S01614

+325   CTTTGTG TGGCTCCTGGAGTCTCCGATCCCAGCCGGACACCCGGGCCTG
      (-)TCF-1 S02023
               (-)gamma-IRE CS S01622
                    (+)gamma-IRE CS S01622
                                 (+)AP-2 S01936
                                 (+)SV40.11 S01003

+374   GTTTCAAA GCGGTCGAACTGCTCTGCCCGCTCCACCGGTAGCGCTCGAG
               (-)PEBP2 S00053
                    (-)CAP-site S00089
                                                        (-)T-Ag S00973

+423   CCTCGGTTT CTCTACTCGACCCCGACTCGCCGCAGCAGGATGATCGCCT
                                          (-)LVc S00040
                                          (-)LVc S00904
                                            (-)CAP-site S00089
                                                    (-)T-Ag S00973

+472   CGCATATGAT CGCCTGCTTATTCACGGAGCTCAACCAAAACCAAGTGCA
        (+)multiple S01614
        (-)multiple S01614
                    (+)LVc S00040
                    (+)LVc S00904
                         (-)AP-1 S01424
                                                (+)WAP US5 S01051
                                                 (+)multiple S01614
                                                 (-)multiple S01614

+521   GAAGGTAAGTC GGCACGGGCGGGAGCTGCTGGCTCGCTTCGGACCAAGT
                        (+)(Sp1) S01028
                        (+)Sp1 S01082
                        (+)Sp1 S00146
                        (+)Sp1 S00800
                        (+)Sp1 S00978
                        (+)JCV repeated se S01193
                        (-)LSF S00975
                        (-)Sp1 S00799
                        (-)Sp1 S00802
                        (-)Sp1 S00803
                        (-)Sp1 S00805
                        (-)EARLY-SEQ1 S01081
                         (+)Sp1 S00781
                         (+)hsp70.2 S00782
                         (-)Sp1 S00801
                         (-)Sp1 S00804
                              (-)CAP-site S00089
                                                (-)H4TF-2 S00742
                                                   (+)WAP US5 S01051
                                                     (-)CAP-site S00089
```

FIGURE 11P

```
+570   TGCGTGCTCTCC GGGAATCTGGAGCACGCAGAGGACCTGCTTCCTCCTC
      (-)XREbf S00395
                        (+)LBP-1 S02121
                        (+)LBP-1 S00487
                               (+)XREbf S00395
                                    (+)polyoma.1 S00922
                                      (+)W-element CS S01629
                                       (+)E-alpha H box S00170
                                       (-)H4TF-2 S00742
                                          (+)LVc S00040
                                          (+)LVc S00904
                                            (-)HC1 S00241
                                             (-)TCF-2-alpha S02024
                                              (-)PEA3 S00392
                                               (-)PU.1 S01502
                                                (+)S1 HS S00506

+619   CGGGGCTGGGGAC GTGGAACCAGTCTGAGTAGCTGGGAAAGTCCTGAGC
       (+)T-Ag S00974
        (-)AP-2 S00346
        (-)AP-2 S01936
        (-)CAP-site. S00089
         (-)AP-2 S01936
           (-)SV40.11 S01003
                          (-)LBP-1 S02121
                          (-)LBP-1 S00487
                           (+)CAP-site S00089
                                (+)AP-1 S02100
                                      (+)H-APF-1 S01971
                                      (-)SV40.11 S01003
                                         (+)AP-3 S01545
                                         (-)IE1.2 S00794
                                            (-)gamma-IRE CS S01622

+668   GCCAGAAACCACGT CTGCTAGGCACCCTCGTGGCCCGGCCGCGCATCAC
       (-)LBP-1 S02121
       (-)LBP-1 S00487
                        (+)T-Ag S00972
                            (+)CAP-site S00089
                                    (+)LF-A1 S00250
                                     (-)GCF S01964
                                         (-)UCE.2 S00437
                                          (+)UCE.2 S00437
                                           (-)GCF S01964

+717   CGATACTCCCACTTT CCCGGGATCCGCGAGCATCCTCCCCACCCTTAAA
               (+)CAP-site S00089
                                       (+)CAP-site S00089
                                             (+)AP-2 S01936
                                             (-)JCV repeated se S01193
                                              (-)PuF S02016
                                                (+)CAP-site S00089

+766   GCCCCTAATTTCTAGA
       (-)T-Ag S00974
                 (+)LBP-1 S00487
                  (-)LBP-1 S00487
```

FIGURE 11Q